

Essays on Health Care Delivery and Financing

by

David C. Chan

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

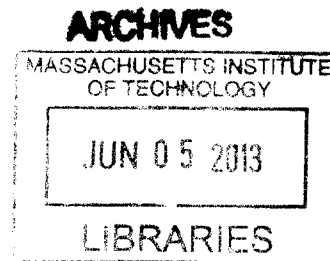
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

© 2013 David C. Chan. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper
and electronic copies of this thesis document in whole or in part.



Author

Department of Economics

May 15, 2013

Certified by.....

A handwritten signature in black ink, appearing to be "Jonathan Gruber".

Jonathan Gruber

Professor of Economics

Thesis Supervisor

Certified by.....

Robert Gibbons

Sloan Distinguished Professor of Management, Sloan School of Management

and Professor of Organizational Economics, Department of Economics

Thesis Supervisor

Accepted by.....

Michael Greenstone

3M Professor of Environmental Economics

Chairman, Departmental Committee on Graduate Studies

Essays on Health Care Delivery and Financing

by
David C. Chan

Submitted to the Department of Economics
on May 15, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis contains essays on health care delivery and financing. Chapter 1 studies the effect of organizational structure on physician behavior. I investigate this by studying emergency department (ED) physicians who work in two organizational systems that differ in the extent of physician autonomy to manage work: a “nurse-managed” system in which physicians are assigned patients by a triage nurse “manager,” and a “self-managed” system in which physicians decide among themselves which patients to treat. I estimate that the self-managed system increases throughput productivity by 10-13%. Essentially all of this net effect can be accounted for by reducing a moral hazard I call “foot-dragging”: Because of asymmetric information between physicians and the triage nurse, physicians delay discharging patients to appear busier and avoid getting new patients.

Chapter 2 explores the development of physician practice styles during training. Although a large literature documents variation in medical spending across areas, relatively little is known about the sources of underlying provider-level variation. I study physicians in training (“housestaff”) at a single institution and measure the dynamics of their spending practice styles. Practice-style variation at least doubles discontinuously as housestaff change informal roles at the end of the first year of training, from “interns” to “residents,” suggesting that physician authority is important for the size of practice-style variation. Although practice styles are in general poorly explained by summary measures of training experiences, rotating to an affiliated community hospital decreases intern spending at the main hospital by more than half, reflecting an important and lasting effect of institutional norms.

Chapter 3, joint with Jonathan Gruber, examines insurance enrollee choices in a “defined contribution exchange,” in which low-income enrollees are responsible for paying for part of the price of insurance. Estimating the price-sensitivity of low-income enrollees for insurance represents a first step for understanding the implications of such a system that will soon become widespread under health care reform. Using data from Massachusetts Commonwealth Care, we find that low-income enrollees are highly sensitive to plan price differentials when initially choosing plans but then exhibit strong inertia once they are in a plan.

Thesis Supervisor: Jonathan Gruber
Title: Professor of Economics

Thesis Supervisor: Robert Gibbons
Title: Sloan Distinguished Professor of Management, Sloan School of Management
and Professor of Organizational Economics, Department of Economics

Acknowledgments

First, I would like to thank my advisers: Jon Gruber, Bob Gibbons, and David Cutler. More than anyone else, they have shaped my development into an economist. They have shown me kindness, humor, patience, and perspective that I hope I can approximate for future trainees I may come across. I also want to specially thank Joe Doyle, who generously provided many rounds of helpful feedback on my work. Numerous other faculty at MIT, too many to list here, have spent valuable time with me and given insightful comments that have improved my research. Together they have created an intellectual environment unmatched anywhere else.

I also would like to thank fellow students I have spent time with at MIT. Some of the brightest yet silliest people I have ever met, they have made my PhD experience incredibly enriching and fun. To name only some, I have benefited from being around Leila Agha, Aviva Aron-Dine, Dan Barron, Nathan Hendren, Brad Larsen, Danielle Li, David Molitor, Christopher Palmer, Iuli Pascu, Mike Powell, Joe Shapiro, Chris Walters, and Tyler Williams. Through mutual support and shared memories, I am fortunate to have them as both friends and colleagues.

I am grateful to many along the way who have influenced me to pursue a PhD in economics. Conversations with Jay Bhattacharya, Bob Brook, Jose Escarce, Alan Garber, David Meltzer, Joe Newhouse, and Susan Ettner left a lasting impression on me, even if on the other end they were just routine chats with a random medical student. David Bates, Joel Katz, and Jane Sillman, from my residency at Brigham and Women's Hospital, could not have been more broad-minded and supportive of my goal to pursue the PhD following residency. Josh Kosowsky and Amy Miller provided key institutional knowledge in the writing of this thesis. I am also grateful for funding from the Agency for Healthcare Quality and Research, the National Institute of Aging, and the Charles King Foundation during my PhD training.

This thesis is dedicated to my family. My parents, Esther and David Sr., tutored me in mathematics and instilled in me a value for delayed gratification, which they probably never thought would lead me down this path. My sister Sara, generous and multi-talented, is a stunning proof-reader of economics, despite being an art historian; she somehow intuitively understands identification and game theory and more importantly knows how to be clear. My wife Helene has always been my greatest supporter; she understands my preferences better than I do. My daughters Chloe and Sophie may be too young now to remember the writing of this thesis, but what they have taken from sleep, they have given back much more in laughter and fulfillment.

Contents

1	Organizational Structure and Moral Hazard among Emergency Physicians	9
1.1	Introduction	9
1.2	Institutional Setting and Data	12
1.2.1	Systems, Pods, and Shifts	12
1.2.2	Physician Schedules and Exogenous Variation	14
1.2.3	Outcomes	15
1.3	Theoretical Framework	15
1.3.1	Stylized Pod Environment	16
1.3.2	Nurse-managed System	17
1.3.3	Self-managed System	20
1.4	Overall Effect of the Self-managed System	21
1.5	Main Evidence of Foot-dragging	24
1.6	Peer Effects on Foot-dragging	27
1.6.1	Presence of a Peer	28
1.6.2	Peer Type	30
1.6.3	Physician and Peer Censuses	31
1.7	Patient Assignment and Equilibrium Building	32
1.8	Conclusion	35
1.A	Empirical Appendix	54
1.A.1	Exogenous Variation in Patients and Physicians	54
1.A.1.1	Similar Exposures across Physician Types	54
1.A.1.2	Joint Insignificance of Physician Identities	54
1.A.1.3	Exogenous Assignment of Physicians to Peers	55
1.A.2	Inference of Overall Effect	56
1.A.2.1	Patient Assignment and Unadjusted Outcomes	56
1.A.2.2	Inference with Serially Correlated Pod-level Error Terms	57
1.A.2.3	Inference with Systematic Placebo Tests (Randomization Inference)	57
1.A.3	Other Behavioral Mechanisms	58
1.A.3.1	Peer Effects	58

1.A.3.2	End of Shift Behavior	60
1.A.4	Additional Results	63
1.B	Theory Appendix	76
1.B.1	Model Setup	76
1.B.2	Nurse-managed System	77
1.B.2.1	Irrelevance of Peer Strategy or Type	78
1.B.2.2	No Physician Reporting, No Triage Nurse Commitment	79
1.B.2.3	No Physician Reporting, Triage Nurse Commitment	80
1.B.2.4	Physician Reporting, Triage Nurse Commitment	81
1.B.2.5	Flow of Expected Future Work	84
1.B.3	Self-managed System	85
1.B.3.1	No Physician Reporting, No Policy Commitment	85
1.B.3.2	No Physician Reporting, Policy Commitment	89
1.B.3.3	Physician Reporting, Policy Commitment	91
1.B.4	Continuous Types	92
1.B.4.1	Intertemporal Feasibility	93
1.B.4.2	Local Incentive Compatibility	94
1.B.4.3	Optimization Problem	95
2	The Dynamics of Housestaff Physician Practice Styles	99
2.1	Introduction	99
2.2	Conceptual Framework	102
2.2.1	Intrinsic Heterogeneity	103
2.2.2	Learning	103
2.2.3	Authority	104
2.3	Institutional Setting	105
2.4	Data	107
2.5	Practice Style Distribution and Dynamics	109
2.5.1	Variation across Physicians	110
2.5.2	Serial Correlation across Time	111
2.5.3	The Role of Continued Practice	113
2.6	Observable Learning Experiences	114
2.6.1	Effect of the Community Hospital	115
2.6.2	Effect of Other Observable Learning Experiences	116
2.7	Intrinsic Heterogeneity	118
2.7.1	Characteristics and Track Choices Prior to Residency	119
2.7.2	Housestaff from Other Residency Programs	120
2.8	Conclusion	121

2.A	Appendix: Tests for Exogenous Assignment	135
2.A.1	Assignment of Patients to Housestaff	135
2.A.2	Assignment of Other Training Experiences to Housestaff	136
3	The Impact of Charging Low-income Families on Health Insurance Choices	145
3.1	Institutional Background of Commonwealth Care	147
3.2	Data and Empirical Strategy	148
3.3	Results	150
3.3.1	Conditional Logit Regressions	150
3.3.2	Predicted Responses to Price Changes	152
3.4	Conclusion	154

Chapter 1

Organizational Structure and Moral Hazard among Emergency Physicians

1.1 Introduction

There is growing recognition that organizational structure could be responsible for setting high-performing health care institutions apart from the rest (McCarthy and Blumenthal, 2006; Oliver, 2007; Institute of Medicine, 2012; Lee and Mongan, 2009). The popular press has noted large differences in cost and quality across health care institutions (Gawande, 2009), and qualitative case studies of the highest-performing institutions have begun to sketch a pattern of organizational attributes that includes concepts such as teamwork, accountability, transparency, and integration (McCarthy and Mueller, 2009).

More broadly, recent research in economics has shown that management and organization can matter greatly in productivity (e.g., Bloom and Van Reenen, 2007).¹ Still, this literature faces challenges in disentangling sources of the organizational effect on productivity. Organizational differences across firms are often difficult to separate from worker selection or other firm-level exposures. Even changes within firms usually involve multiple features, which complicates isolating behavioral mechanisms. With respect to mechanisms, a gap in understanding remains between the classic prediction that workers in teams should engage in moral hazard (Holmstrom, 1982) and findings in the management literature, largely outside of economics, suggesting that workers perform better with teamwork (e.g., Yeatts and Hyten, 1997).²

¹Other landmark studies include Ichniowski et al. (1997); Bertrand et al. (2004); Bloom et al. (2011); Hamilton et al. (2003).

²Teamwork has been most extensively studied in the management and organizational behavior literature, usually with case studies or cross-sectional studies. Explanations have been largely psychological, including concepts such as “cohesion,” “recognition,” and “motivation” (Yeatts and Hyten, 1997). Hamilton et al. (2003), a study in the economics literature, posited complementary skills between team members, found some evidence of selection of high-productivity workers into teams, and concluded that these effects must have outweighed increased moral hazard in teams.

I study a natural experiment in which emergency department (ED) physicians are observed to work in two different organizational systems that differ in only one respect. In one system, which I call “nurse-managed,” two physician “workers” in the same location (“pod”) are individually assigned patients by a triage nurse “manager.” In the second system, which I call “self-managed,” the triage nurse first assigns patients to a pod that is shared by two physicians, and then these physicians decide between themselves who will care for each arriving patient. These two systems are common in ED settings (Salluzzo et al., 1997; Patel and Vinson, 2005). More generally, this comparison highlights an important organizational dimension: the degree of autonomy workers have to manage and coordinate work among themselves.

Outcomes in the two systems may differ via several mechanisms. In the nurse-managed system, under asymmetric information between physician workers and the triage nurse manager, physicians may want to avoid being assigned more work by appearing busier than they are, keeping patients longer than necessary (“foot-dragging”). In the self-managed system, if physicians have more information on each other’s true workloads and use it in choosing patients, they can reduce foot-dragging relative to the nurse-managed system. However, physicians in the self-managed system may also seek to avoid work by waiting for their peer to pick patients first (another moral hazard that I distinguish with the term “free-riding”). Finally, outcomes may differ through advantageous selection in the self-managed system, as physicians can choose patients according to either skill or availability.

Several features of this empirical design allow for identification both of the overall effect of self-managed teams and of foot-dragging and free-riding as specific mechanisms. First, one of the two pods changed from a nurse-managed system to a self-managed system during the sample period, allowing me to control for time-invariant and unobservable differences between the two pods. Second, I observe the same health care providers – physicians, nurses, and residents – working in both pods. Third, physician schedules are arranged far in advance and do not allow physicians to choose shifts precisely. I construct a measure based on the exogenous flow of work to the ED to isolate foot-dragging, and I use exogenous variation in the assignment of peers to evaluate peer effects on foot-dragging. Fourth, the detailed nature of physician orders allows me to infer when physicians start working on a case and to isolate free-riding.

I focus on the time a physician spends on a patient (i.e., the patient’s length of stay) as my primary outcome measure in the ED. This resembles measures of throughput used in other studies of worker productivity (Mas and Moretti, 2009; Bandiera et al., 2009, 2010). In the ED, throughput is especially relevant because it impacts waiting times, a key determinant of patient satisfaction and outcomes (Bernstein et al., 2008; Thompson et al., 1996). Waiting times are shared by all future patients and depend on aggregate ED throughput, while lengths of stay map each physician’s contribution to this aggregate. Consistent with the health care setting, I also examine secondary patient-level outputs of quality, revenue, and costs. I do not expect them to differ much if pure foot-dragging moral hazard is the primary mechanism and if the main concern for quality arises from increased waiting times shared by all future patients.

I find that physicians perform 10-13% faster in the self-managed system than in the nurse-managed system. There is no difference in the time physicians take to write their first orders for each patient, which suggests that free-riding is not a significant mechanism. I then examine foot-dragging by testing a prediction that, under this mechanism but not other mechanisms, lengths of stay should increase with *expected future work*. I find that lengths of stay increase with expected future work in the nurse-managed system but not in the self-managed system. This essentially explains the overall difference between the two systems.

Further, I show that information between peers and social incentives play a crucial role in modifying foot-dragging. To test this, I use the fact that the location of other physicians does not affect total work but could affect the ability of physicians to monitor each other's work. Thus, if physicians care about being seen engaging in moral hazard, the presence of a peer in the same pod could reduce foot-dragging. I find that the presence of a peer in the same pod – especially a more senior peer – substantially reduces foot-dragging in the nurse-managed system, relative to when there is another physician in the ED but not in the same pod. I also find that foot-dragging in the nurse-managed system depends on the number of patients (the “census”) of both the index physician and the peer, reflecting strategic behavior or social incentives conditional on the distribution of work.

Finally, I study patient assignment to test a prediction reminiscent of Milgrom and Roberts (1988): If the triage nurse knows that physicians are tempted to foot-drag (i.e., distort their censuses upward as signals), then she can be better off by ignoring some information in censuses. That is, she can improve patient assignment *ex ante* by committing to a policy that ignores informative signals *ex post*. In contrast, in the self-managed system, if physicians observe and use information about each other's true workloads, then signals are less likely to be distorted and can be used more efficiently *ex post*. Consistent with this, I find that patient assignment is more correlated with censuses in the self-managed system than in the nurse-managed system. I also study patient assignment around the transition of the pod switching to a self-managed system and find evidence of short-term enforcement against foot-dragging, with physicians with higher censuses *more* likely to be assigned new patients, prior to improving *ex post* efficiency.

Together, these findings suggest that the self-managed system improves performance because superior information between peers is used to assign work, reducing the moral hazard to avoid work. The remainder of the paper proceeds as follows. The next section describes the ED institutional setting and data. Section 1.3 outlines a simple model of asymmetric information in the nurse-managed system to explain foot-dragging there and its reduction in the self-managed system; it also discusses conditions for free-riding in the self-managed system. Section 1.4 reports the overall effect of the self-managed system and shows that free-riding is minimal in the self-managed system. Sections 1.5 and 1.6 discuss the main evidence for foot-dragging and its mitigation by organizational structure and the presence of peers. Section 1.7 explores patient assignment in the two systems over time. Section 1.8 concludes.

1.2 Institutional Setting and Data

I study a large, academic, tertiary-care ED with a high frequency of patient visits, greater than 60,000 visits per year (or 165 visits per day), with a total of 380,699 visits over six years. For each visit, I observe times for each point in the process of care – patient arrival at the ED, arrival at the pod, entry of discharge order, and discharge with destination – as well as all physician orders written during the visit (approximately 13 orders on average per visit). Because all actions taken by physicians must be documented and time-stamped as orders, these data provide uniquely detailed process measures of physician effort and patient care.

Patient care in the ED is delivered by an attending physician (“physician”), a nurse (not to be confused with the triage nurse), and sometimes a resident physician (“resident”). The physician is responsible for directing patient care, while the resident, who is still training, may assist the physician with varying levels of autonomy. Nurses execute physician orders and report any concerns to physicians. I observe 92 physicians, 364 nurses, and 986 residents in the data. Among these, 75 physicians, 334 nurses, and 882 residents, comprising 11,865 unique physician-nurse-resident trios, are observed in both organizational systems.³

1.2.1 Systems, Pods, and Shifts

All patients must first enter the ED through a waiting room, also called “triage,” where a triage nurse decides where and when to send them. In the nurse-managed system, the triage nurse directly assigns patients to one of potentially two physicians in the same pod. The triage nurse serves as a manager in the sense that she allocates new patients to physicians whom she thinks are available or able to do the work.⁴ In the self-managed system, two physicians in the same pod are jointly responsible for dividing work sent to the pod by the triage nurse. See Figure 1.2.1 for a schematic of patient workflow. The assignment of patients to nurses and residents does not differ between the two systems; in both systems, nurses are assigned patients, and residents choose patients.

Basic information about patients cared for by each physician is available to all physicians in the ED and the triage nurse from a computer interface (see Figures 1.2.2 and 1.2.3). The most important public measure of workload is each physician’s census, or the number of patients being cared for by him. However, censuses can be distorted by prolonging the time to discharge. Physicians may have superior information relative to the triage nurse about the true workloads of their peers, not only from differences in medical knowledge but also because they are in the same room and can directly observe peer behavior and patient status. In the self-managed system, physicians may also use this

³Essentially all providers who do not work in both systems either are occasional moonlighters or represent errors in recording the correct provider. For example, the number of visits corresponding to median resident is 1,525, while this number is 17 for residents who are observed to work in only one system.

⁴The assignment of patients by nurses or non-medical staff is the predominant system of work assignment in hospital and ED settings. In many of these settings, however, these “managers” have no discretion but merely follow rules. Neither, of course, can they hire or fire nor set financial incentives.

information to assign patients.

Physicians in the ED work in prescheduled shifts of eight to nine hours. Each shift is in one of two geographic locations, or pods, which I call “Alpha” and “Bravo.” Alpha and Bravo pods are similar in resources, layout, and staffing, and they have remained so over time. The triage nurse can decide to send any patient to either pod, based on bed availability. However, one important difference between Alpha and Bravo is that Alpha pod has always been opened 24 hours, while Bravo pod has always closed at night. As a result, patients who need to stay longer, either because they are sicker or have conditions that might make discharge difficult (e.g., psychiatric patients), have tended to be sent to Alpha pod.⁵ Closing Bravo every night may also prompt earlier discharges for patients in the pod as it nears closing.

Alpha pod has always had a self-managed system. In contrast, in March 2010, Bravo pod switched from a nurse-managed system to a self-managed one. The regime change in Bravo pod resulted from a simple intervention in which beds that physicians previously “owned” became shared, so that the physicians were then allowed to choose among patients entering that pod. The reason for this switch, according to the ED administration, was to allow greater flexibility in patient assignment within pod. According to interviews with ED administrators and physicians, the switch was not considered a significant change in organizational structure, and overall implications for efficiency were not apparent.⁶

Importantly, there was no other official change in either Alpha or Bravo accompanying the above regime change: Schedules and staffing for providers and algorithms for patient assignment to beds, nurses, and residents remained unchanged. Actual assignment choices by the triage nurse between the pods were also relatively stable over time; if anything, Bravo pod received increasingly time-intensive patients over time, as the ED became increasingly busy.⁷ Financial incentives for physicians were unchanged; they have always been paid a salary plus a 10% productivity bonus based on clinical productivity (measured by Relative Value Units, or RVUs, per hour) and modified by research, teaching, and administrative metrics.⁸

⁵For example, the average patient age in Alpha was about 50, while the average patient age in Bravo was about 45. The average patient Emergency Severity Index was 0.5 points more severe (on a 5-point scale, with lower numbers being more severe) for patients in Alpha compared to those in Bravo. These and other summary statistics comparing patients in the two pods are shown Table 1.A.2.

⁶In fact, in May 2011, the ED attempted a redesign that moved both pods to the nurse-managed system, only to discover later that it significantly reduced efficiency. They reversed this organizational change in January 2012.

⁷In Section 1.A.2.1, I calculate the expected length of stay for patients based only on their characteristics. I then average the expected lengths of stay for each pod and each month. In the first month of the data, patients in Alpha had characteristics that predicted a 14% greater length of stay than those for patients in Bravo; in the last month of the data, this incremental percentage was only 6%. I show this graphically in Figure 1.A.2.

⁸The metric of Relative Value Units (RVUs) per hour encourages physicians to work faster, because RVUs are mostly increased on the extensive margin of seeing more patients and are rarely increased by doing more things for the same patients. I specifically address whether physicians can bill for more RVUs in the self-managed system or when foot-dragging in Sections 1.4 and 1.5, respectively.

1.2.2 Physician Schedules and Exogenous Variation

In addition to the natural experiment provided by the regime change, this study exploits other identifying variation due to physician shift assignment. Physician schedules are determined one year in advance, and physicians are only able to request rare specific shifts off, such as holidays or vacation days.⁹ General preferences, such as whether they would like to work at night, may be voluntarily stated but not honored fully, and all physicians are expected to be open for shifts at all times of the day and days of the week. Once working on a shift, physicians cannot control the volume and types of patients arriving in the ED nor the types of patients assigned by the triage nurse to the pod.¹⁰

Conditional on the month-year, day of the week, and hour of the day, I find that physicians are exposed to similar patients types arriving at their pod and patient numbers arriving at the ED. Tables 1.2.1 and 1.2.2 show this descriptively for physicians with above- and below-median productivity, defined by average lengths of stay. In Section 1.A.1, I show similar results for physicians with different preferences for certain patients (defined as the likelihood to choose that patient when in the self-managed system); I formally test and cannot reject the null that physician identities are jointly insignificant in predicting available patient types or ED arrival numbers; and I show that physician types are uncorrelated with those of their peers.¹¹

The observed variation is not only exogenous but also rich for several reasons: First, the numbers and types of patients arriving at the ED are notoriously wide-ranging, even conditional on the time of the day.¹² Second, physicians work very few shifts per week, usually one to two with the maximum being four, and are expected to work in all types of shifts. As a result, I observe all physicians working in both locations, during all time categories, and with essentially all possible peers. Third, there is substantial variation in the tenure of physicians. While some physicians are observed to remain on staff for the entire six-year period, other physicians are newly hired or leave the hospital during the observation period. I observe physician demographics and employment details, such as the place and date of medical school and residency. I use these data to construct rich descriptors

⁹Shift trades are also exceedingly rare, about less than one per month, or <1% of the number of shifts. Results are robust to dropping traded shifts. Per ED administration, shifts are not assigned with peers in mind.

¹⁰Physicians may rarely (<1-2% of operating times) put the ED on “divert” for up to an hour when the flow of patients is unusually high and the entire ED lacks capacity to see more patients. Even when this happens, this only affects some ambulances (which as a whole constitute 15% of visits) carrying serious emergencies, as opposed to the majority of patients, some of whom walk in. ED flow is largely unaffected.

¹¹Note that I already can control for physician identity across systems. So even if physicians are preferentially assigned certain shifts (even after conditional on rough time categories), estimates of the overall self-managed effect and of foot-dragging across systems will still be unbiased as long as conditions associated with those shift times have the *same* effect on each physician’s outcomes or foot-dragging, respectively, in both systems. Exogenous physician assignment provides additional robustness allowing for differential effects of unobservable conditions within physicians across systems. Regression specifications are discussed further below.

¹²At any given hour, the number of patients arriving may range from close to none to the mid-twenties. Patients may require a simple prescription or pregnancy test, or they may have a gunshot wound.

of peer relationships, described in Section 1.6.2.

1.2.3 Outcomes

ED length of stay is my primary outcome measure because it directly relates to ED throughput productivity, consistent with other studies of worker productivity (e.g., Mas and Moretti, 2009). Throughput is especially important in the ED, as the focus of both policy papers and a cottage industry of ED management consulting (McHugh et al., 2011). Given ED bed capacity constraints and patients almost always in waiting rooms, lengths of stay determine waiting times, believed to be important for patient satisfaction and health outcomes (Thompson et al., 1996; Bernstein et al., 2008). Lengths of stay measure each physician’s contribution to aggregate waiting times. Measured from the arrival at the pod to entry of the discharge order, they are unaffected by inpatient bed availability, patient home transportation, or clinical care or patient adherence after ED visits.

I consider three secondary outcome measures of quality that have been prominent at the national policy level (Schoor and Venkatesh, 2012; Forster et al., 2003; Lerman and Kobernick, 1987). Thirty-day mortality occurs in about 2% of the sample. Hospital admission represents a resource-intensive option for ED discharge that is believed to substitute sometimes for appropriate care in the ED and occurs in 25% of the sample. Bounce-backs, defined as patients who are discharged home but return to the same ED within 14 days, occur in about 7% of the sample and represent a complementary quality issue.

I also consider patient-level revenue and costs that accrue to the ED and hospital. For revenue, I use Relative Value Units (RVUs), which are units of physician billing for services that scale directly to dollars and reflect the intensity of care provided to a patient.¹³ For costs, I use total direct costs for each patient encounter, including any costs incurred from a resulting hospital admission.¹⁴ Finally, because I have data on all orders, I consider detailed process measures that capture all aspects of patient care, including nursing, medication, laboratory, and radiology orders. I do not observe the time that a physician officially signs up for a patient, but as a proxy for this, I use the time that the physician writes his first order.

1.3 Theoretical Framework

In this section, I outline a simple model of asymmetric information between physicians and the triage nurse. The purpose of this model is to show how the self-managed system reduces foot-dragging and improves assignment efficiency relative to the nurse-managed system, formalizing the conjecture in the management literature that self-managed teams improve productivity by “monitoring and

¹³The current “conversion factor” is \$34 per RVU, and the average ED patient is billed for 2.7 RVUs of ED care, resulting in about \$6 million in yearly revenue for this particular ED.

¹⁴Direct costs are for services that physicians control and are directly related to patient care. Indirect costs include administrative costs (e.g., paying non-clinical staff, rent, depreciation, and overhead).

managing work process and progress” (Pallak and Perloff, 1986).

I assume that in the nurse-managed system, the triage nurse cannot observe true physician workloads, although she would like to assign new work according to workloads. Given that physicians prefer to avoid new work, they distort signals of true workload by prolonging patient lengths of stay (i.e., foot-dragging). At the same time, similar to Milgrom and Roberts (1988), I show that a triage nurse who takes this into account can be better off by committing to an *ex post* inefficient policy of ignoring signals, even though signals remain informative.

In the self-managed system, however, if physician peers sometimes observe each other’s true workloads, then they can also use that information to assign new work. This reduces the threat of foot-dragging and improves *ex post* assignment efficiency. While physicians may delay choosing patients in the self-managed system (i.e., free-riding) as another moral hazard to avoid work, this is limited with sufficient mutual physician information or commitment. Finally, I contrast self-management with social incentives as another pathway for information between peers to reduce foot-dragging: Social incentives reduce foot-dragging simply because physicians do not want to be seen engaging in it (e.g., Kandel and Lazear, 1992; Mas and Moretti, 2009).

1.3.1 Stylized Pod Environment

Consider the following simple game of asymmetric information: Two physicians $j \in \{1, 2\}$ work in a single pod at the same time. They each have one patient, entailing a low or high workload. In addition to the time that they take on their current patients, physicians also care about new work – a third patient – that might be assigned to one of them. Physician utility is given by

$$u_j^P = -(t_j - \theta_j)^2 - K_P(\theta_j) \mathbb{I}\{\mathcal{J}(3) = j\}, \quad (1.3.1)$$

where t_j is the time that physician j keeps his initial patient, $\theta_j \in \{\underline{\theta}, \bar{\theta}\}$ is the workload entailed by his initial patient (where $\bar{\theta} > \underline{\theta} > 0$), $K_P(\theta_j) > 0$ is the cost of getting a potential third patient conditional on θ_j , and $\mathcal{J}(3)$ denotes the physician who gets the third patient.

Type $\underline{\theta}$ occurs with probability p , and type $\bar{\theta}$ occur with probability $1 - p$. Types are never observed by the triage nurse, but with probability ψ , peers can observe each other’s initial patient types. In contrast, the number of patients of each physician (his census, $c_j \in \{0, 1\}$) is public information at any time. The action that each physician takes is t_j . Absent any strategic behavior, each physician would like to discharge his patient at $t_j = \theta_j$, which I also assume is the socially optimal time for treatment and generically captures all concerns of care (e.g., patient health and satisfaction, malpractice concerns, physician effort and boredom).¹⁵ Because patient assignment depends on the organizational system, I discuss the triage nurse and assignment further in the following subsections.

¹⁵I thus abstract away from any principal-agent problem between physician and patient or between physician and ED in this term, other than the cost incurred by the new patient, discussed below.

The physician assigned the new patient incurs a cost, which depends on his initial workload, θ_j . I specify this cost as $K_P(\underline{\theta}) = \underline{K}_P$ and $K_P(\bar{\theta}) = \bar{K}_P$, where $\bar{K}_P > \underline{K}_P > 0$. This reflects the idea that neither physician would like to get the new patient, but that it is more costly for a busy physician,¹⁶ either because he must exert more additional effort or because he will have worse outcomes for this new patient.

The timing of the game is as follows:

1. At time $t = 0$ physicians each receive one patient, discovering $\theta_j \in \{\underline{\theta}, \bar{\theta}\}$.
2. Physicians simultaneously choose how long they will keep their patients, t_j .
3. With probability $\psi > 0$, physicians observe each other's θ_j .
4. Exactly one patient will arrive with uniform probability distributed across the time interval $t \in [\underline{\theta}, \bar{\theta}]$. When the patient arrives, this new patient is assigned to a physician by the triage nurse (in the nurse-managed system) or the physicians themselves (in the self-managed system).
5. Physicians complete their work on the one or two patients under their care and end their shifts. They receive payoffs given in Equation (1.3.1).

This model highlights the tension between using signals (censuses c_j) of private information (types θ_j) for patient assignment and the fact that these signals can be distorted (through t_j). Regardless of their type, physicians prefer to avoid new work (through $K_P(\theta) > 0$), but otherwise I assume that physicians have no incentive to keep patients longer than socially optimal. Although the triage nurse never observes the physicians' types, physicians observe each other's types with probability ψ . This superior information allows greater efficiency through two separate pathways – self-management and social incentives – discussed below.

1.3.2 Nurse-managed System

In the nurse-managed system, the triage nurse assigns the new patient to a physician. In my baseline model, I assume that physicians cannot report their types or anything else to the triage nurse, but that the triage nurse can credibly commit to an assignment policy prior to physicians receiving their patients. I believe that this scenario is most realistic: Physicians will find it difficult to describe their workloads in practice, but an assignment policy based on censuses is easily observable, relatively simple, and can be enforced in a repeated game.¹⁷

¹⁶To be precise, a “busy” physician is a physician who starts with a high-workload patient (i.e., his type is $\bar{\theta}$). Merely keeping his patient longer, prolonging t_j , does not make a physician busier. This interpretation is supported by evidence that physicians do not provide more care for patients they are foot-dragging on, shown in Section 1.5.

¹⁷In Appendix Section 1.B.2, I consider two alternative scenarios: (1) the pure signaling game which allows neither physicians to report their types nor the triage nurse to commit to an assignment policy, and (2) the mechanism design game (without transfers) which allows physicians to report types and the triage nurse to commit to a policy. Results are similar, but not surprisingly, reporting and commitment as additional capabilities both improve efficiency.

The triage nurse's utility is

$$u^N = -D \sum_{j \in \{1,2\}} (t_j - \theta_j)^2 - K_N(\theta_{\mathcal{J}(3)}), \quad (1.3.2)$$

which is similar to but potentially different from that of the physicians. D is an indicator that allows the triage nurse to care about the treatment times of the first two patients as outcomes (if $D = 1$). Remember that the socially optimal discharge times for patients is $t_j = \theta_j$, which is universally agreed upon.

The second term, $K_N(\theta)$, is the cost of assigning the new patient to a physician whose initial patient is of type θ , specified as $K_N(\underline{\theta}) = 0$ and $K_N(\bar{\theta}) = \bar{K}_N$, where $\bar{K}_N > 0$. As before, this represents some relative cost (now to the triage nurse) in assigning the new patient to a physician with greater workload, because that physician will take more time or provide lower quality of care. I do not restrict the value of \bar{K}_N relative to $\bar{K}_P - \underline{K}_P$.¹⁸

At $t = 0$, the triage nurse commits to an assignment policy function $\pi(c_1, c_2)$, with censuses $c_j \in \{0, 1\}$. To simplify the analysis, I impose a symmetric policy function with $\pi(0, 0) = \pi(1, 1) = \frac{1}{2}$ and $\pi \equiv \pi(0, 1) = 1 - \pi(1, 0)$. That is, when both physicians have equal censuses, the triage nurse should have no preference to send the new patient to one physician or the other, since she has no other information about who is less busy at that time. Note that $\pi = 1$ represents what I mean by *ex post* efficiency, since the triage nurse can infer that if $c_j = 0$ and $c_{-j} = 1$, then j certainly had the lower workload.

I use a Perfect Bayesian Equilibrium as the equilibrium concept. In equilibrium, the triage nurse chooses the optimal assignment policy, summarized by $\pi^* \equiv \pi^*(0, 1)$, given physician discharge strategies \underline{t}^* and \bar{t}^* for initial patients of type $\underline{\theta}$ and $\bar{\theta}$, respectively. Given π^* , physicians choose optimal discharge strategies \underline{t}^* and \bar{t}^* .

Proposition 1. *In the Perfect Bayesian Equilibrium for the nurse-managed system, physicians with $\underline{\theta}$ and $\bar{\theta}$ discharge their patients at $\underline{t}^* > \underline{\theta}$ and $\bar{t}^* = \bar{\theta}$, respectively, and the triage nurse assigns the new patient to the physician with census 0, when the other physician has census 1, with some probability π^* between $\frac{1}{2}$ and 1.*

First note that the triage nurse will never want to send the new patient with greater probability to a physician with $c_j > c_{-j}$. So physicians with high-type patients will never want to mimic those with low-type patients and will choose $\bar{t}^* = \bar{\theta}$. But physicians with low-type patients have some reason to mimic having high-type ones. For a given $\pi \equiv \pi(0, 1)$ previously chosen by the triage nurse, the optimization problem of physicians with $\underline{\theta}$, $\max_{t_j} \mathbb{E} \left[u_j^P(t_j; \pi, \underline{\theta}) \right]$, yields the first-order condition

$$\underline{t}^* = \underline{\theta} + \frac{\underline{K}_P}{2(\bar{\theta} - \underline{\theta})} \left(\pi - \frac{1}{2} \right). \quad (1.3.3)$$

¹⁸Note also that if $D = 0$, then it does not matter what value \bar{K}_N takes, as long as it is some positive number.

For physicians with $\underline{\theta}$, there is a first-order gain in temporary mimicry of having $\bar{\theta}$ relative to a second-order loss, as long as $\pi > \frac{1}{2}$. In other words, as long as the triage nurse is more likely to send the new patient to a physician she believes is less busy, physicians with $\underline{\theta}$ will foot-drag to mimic those with $\bar{\theta}$.¹⁹

The triage nurse commits to π^* that maximizes her expected utility given \underline{t}^* and \bar{t}^* . Substituting (1.3.3) into her expected utility and solving the first-order condition yields π^* . For simplicity, I present the expression for π^* if $D = 0$:

$$\pi^* = \frac{1}{2} + \frac{(\bar{\theta} - \underline{\theta})^2}{K_P}. \quad (1.3.4)$$

Equation (1.3.4) shows that the nurse's choice of π depends on the cost of getting the new patient for the *physician with lower workload*, because of his temptation to foot-drag. As this temptation increases, it is optimal for the triage nurse to decrease *ex post* assignment efficiency, π^* . In Appendix Section 1.B.2.3, I show that π^* is even lower if $D = 1$: If the triage nurse also cares about lengths of stay, which is more realistic, then the triage nurse puts even less weight on the physicians' censuses when assigning the new patient.

The important general point is that the triage nurse may commit to an *ex post* inefficient assignment policy function, $\pi^* < 1$. Even if she wants to optimize only the assignment of the third patient, this policy may improve her expected utility, which is similar to Milgrom and Roberts' (1988) result that managers can be better off if they commit not to listen to subordinates who could undertake costly "influence activities." This commitment increases triage-nurse utility by decreasing foot-dragging, shown in Appendix Section 1.B.2.

The assumption of a single patient arriving in the interval $t \in [\underline{\theta}, \bar{\theta}]$ is convenient for representing the temptation of moral hazard for physicians with $\underline{\theta}$. However, in practice there are of course many new patients, and I identify foot-dragging as the response of lengths of stay to the flow of expected future work, defined in terms of *numbers* of patients arriving at the ED triage. To capture this intuition, I can extend the model by changing the interval over which the single patient is expected to arrive, which results in replacing the interval $\bar{\theta} - \underline{\theta}$ in the denominator of Equation (1.3.3) with some $\Delta t \leq \bar{\theta} - \underline{\theta}$, as long as \underline{t}^* is an interior solution. I show details in Appendix Section 1.B.2.5, but the intuition is straightforward. With an infinite flow of patients to the ED (as $\Delta t \rightarrow 0$), physicians should expect to get a new patient the minute they discharge one. With no expected future patients (as $\Delta t \rightarrow \infty$), there is no incentive to foot-drag.

¹⁹A nice feature of this simple two-type model is that the first-order condition does not depend on what the peer's type or strategy is, because there is only one patient each and thus one policy parameter π . Keeping this initial patient longer by dt decreases the likelihood of getting the new patient by $(\pi - 1/2) dt / (\bar{\theta} - \underline{\theta})$ regardless of the peer's census. For this reason, parameters like p do not matter. This is shown in detail in Appendix Section 1.B.2.1.

1.3.3 Self-managed System

For the self-managed system, I assume the same physician utilities and information structure as in the nurse-managed system. The only difference is that the two physicians, not a triage nurse, are responsible for deciding who gets the new patient. Physicians choose both t_j and an action that determines the assignment of the new patient. I consider two microfoundations of this assignment action in the self-managed system, both continuing the baseline assumption that physicians cannot report their types.²⁰ In one microfoundation, each physician may only decide whether to choose an unattended patient at each point in time. Alternatively, physicians may commit to an assignment policy that uses not only censuses but also observations of true workload, which are available with probability ψ . I present a brief discussion of results below; details and derivation are in Appendix Section 1.B.3.

Proposition 2. *Consider the Perfect Bayesian Equilibrium for the self-managed system. If physicians cannot commit to a policy function, there will be no foot-dragging or ex post inefficient assignment, but some free-riding if $\psi < 1$. If physicians can commit to a policy function, and if $\psi > 0$, $\bar{K}_P - \underline{K}_P > \bar{K}_N$, and $D = 1$, then there will be less foot-dragging and more ex post efficient assignment, relative to the nurse-managed system, and no free-riding.*

Without physician commitment to an assignment policy, physicians play a war of attrition (e.g., Bliss and Nalebuff, 1984), both incurring a cost by the new patient remaining unattended after arrival, which I specify in detail in Appendix Section 1.B.3.1. If types are observed and different, the physician with $\underline{\theta}$ will choose the new patient immediately upon arrival by subgame-perfect reasoning similar to Rubinstein (1982). There is no point in waiting if he knows that his peer will wait longer (because the peer with $\bar{\theta}$ has a higher cost of accepting the new patient). On the other hand, if they do not observe each other's types, or if they observe that they have the same type, then they engage in free-riding, inefficiently leaving the new patient unattended with some probability. Physicians with $\bar{\theta}$ refrain from choosing the new patient with any probability if there is a chance their peer has $\underline{\theta}$. Thus, there is no foot-dragging by physicians with $\underline{\theta}$, because it never reduces the chances of getting the new patient.²¹

In the second microfoundation, physicians commit to an assignment policy similar to the triage nurse's assignment commitment in the nurse-managed system. However, with probability ψ , physicians can use each other's observed types to assign the new patient, and in this event, foot-dragging will have no effect on assignment. This chance, however small, lowers the attractiveness of foot-dragging. With probability $1 - \psi$, physicians follow an assignment policy chosen to maximize their

²⁰In Appendix Section 1.B.3.3, I consider the case in which physicians can report their types to each other. This mirrors the case of physician reporting in the nurse-managed system, in Appendix Section 1.B.2.4. In both cases, reporting improves efficiency.

²¹At the end of Appendix Section 1.B.3.1, I discuss that there is some foot-dragging with continuous types, but that foot-dragging will still be less than in the nurse-managed system with no triage nurse commitment, in Appendix Section 1.B.2.3.

ex ante expected utility with the knowledge that the new patient is more costly for a busy physician but that, upon discovering they are less busy, they will still foot-drag. They can afford to commit to an assignment policy with greater *ex post* efficiency than the triage nurse’s in the nurse-managed system, primarily because foot-dragging is less of a threat with more information on true workloads.²²

In summary, under both microfoundations (with reasonable parameters in the commitment case), physicians foot-drag less, and the new patient is given more often to a physician with lower workload. I consider both microfoundations because the truth likely lies in between: Physicians are not forced to see patients in the self-managed system, but they very likely have a strong cultural norm that quickly assigns patients in order to prevent patients from waiting in the pod unattended for too long. The model also predicts little free-riding with sufficient physician information (high ψ), commitment, or costs of leaving patients unattended.

The key point is that information between physicians reduces the usefulness of moral hazard to avoid new work and improves assignment efficiency because physician actions determine assignment. It is useful to contrast this with social incentives, which also reduce moral hazard by information between peers. A growing empirical literature suggests that social incentives operate because peers incur a social cost when seen engaging in moral hazard (Bandiera et al., 2005, 2009; Mas and Moretti, 2009; Jackson and Schneider, 2010).²³ Note that self-management can improve efficiency without social incentives and that social incentives may operate in the nurse-managed system. Nevertheless, because self-management and social incentives both depend on superior information between peers, peer effects on foot-dragging, examined in Section 1.6, are a useful test of both social costs and superior information between peers ($\psi > 0$).

1.4 Overall Effect of the Self-managed System

In this section, I estimate the overall effect of the self-managed system on a given team of providers and on a given patient. I specifically ask the following: If the same patient and providers were assigned to each other in a different organizational system, what would their outcomes be? I can control for pod-specific time-invariant unobservable differences by the fact that I observe one of the two pods (Bravo) switching from a nurse-managed system to a self-managed one. I can also control for providers because I observe essentially all providers – physicians, residents, and nurses – working in both pods over time. Of course, I do not observe the same patient visit in both pods, but I can

²²In Section 1.B.3.2, I show that another reason for improved *ex post* assignment inefficiency is that physicians are likely to care relatively more about inefficient assignment than the triage nurse (i.e., $\Delta K_P > \bar{K}_N$), since the cost of inefficient assignment is scaled relative to treating their own patients, while the triage nurse scales this relative to treating all patients.

²³Social incentives have also been considered theoretically (Kandel and Lazear, 1992). It is straightforward to modify physician utility in (1.3.1) such that their expected utility includes a term $\psi S(\cdot)$, where $S(\cdot)$ are social costs incurred conditional on being seen foot-dragging, that reduces foot-dragging.

condition on a rich set of patient characteristics.

As my baseline analysis, I estimate the following equation:

$$Y_{ijkpt} = \alpha \text{Self}_{pt} + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.4.1)$$

where outcome Y_{ijkpt} is indexed for patient i , physician j , resident-nurse k , pod p , and arrival time t . The variable of interest in Equation (1.4.1) is Self_{pt} , which indicates whether pod p had a self-managed system at time t . It also controls for patient characteristics X_{it} , pod identities ζ_p , a sum of fixed effects for time categories η_t (for month-year, day of the week, and hour of the day), and physician-resident-nurse trio identities ν_{jk} . By controlling for observable patient characteristics and provider identities, I only require parallel trends in outcomes conditional on this additional information (Abadie, 2005).²⁴

Because I cannot control for patient unobservables, I assume that average unobserved characteristics of patients sent to Alpha versus Bravo did not change over time. In Table 1.4.1, I estimate several versions of (1.4.1), including progressively more controls for patient characteristics. The estimate for the effect of self-managed teams on log length of stay remains stable (and slightly *increases* in magnitude) from -10% to -13% upon adding a progressively rich set of controls. This is consistent with the fact that sicker patients were sent to Bravo pod over time (see Appendix Figure 1.A.2) and suggests that unobserved characteristics may follow the same pattern.

This overall effect represents a significant difference in length of stay due to a simple organizational change in which physicians assign work among themselves, while the physicians themselves and financial incentives were held fixed. As a comparison, this effect is roughly equivalent to one standard deviation in physician productivity fixed effects: Physicians who are one standard deviation faster than average have lengths of stay that are about 11% shorter. Given average lengths of stay, the self-managed-system effect is equivalent to a reduction in lengths of stay by 20-25 minutes per patient, and under simple assumptions, it represents a \$570,000 yearly savings to this single ED.²⁵

While I find a significant effect of self-managed teams on length of stay, I find no statistically significant effect on quality outcomes (30-day mortality, hospital admissions, 14-day bounce-backs) and financial/utilization outcomes (RVUs and total direct costs), shown in Table 1.4.2.²⁶ Alternative

²⁴As discussed in Section 1.2, there were secular trends between the two pods. Specifically, more and sicker patients were sent to Bravo pod over time, and new nurses generally spent more time in Bravo pod to fill the need of higher volume. I show unconditional results and discuss this in Section 1.A.2.

²⁵For this back-of-the-envelope calculation, I simply assume that ED patient volume is exogenous and that the ED is able to reduce the number of physician-hours, given improved throughput, to meet the volume. By allowing more patients to be seen for a given number of physician-hours, the \$570,000 yearly saving to this ED derives from \$4.4 million yearly spending in physician hourly salaries (26,280 physician-hours per year at about \$167 per physician-hour). This gain ignores reduced waiting times and improved outcomes shared by all ED patients.

²⁶While I can rule out 0.8% increase in mortality, this is relatively large compared to the baseline 2.0% mortality. Effects are more precisely estimated for other outcomes, and I can rule out 4% relative increases in revenue and costs.

mechanisms of free-riding and advantageous selection could affect the quality of care and utilization, because they mean that specific patients either are being made to wait for care or are seen by physicians who are better suited (or more available) to see them. In contrast, under pure foot-dragging, only the *discharge* of patients is delayed in order to prevent more work. Foot-dragging should not result in different quality or utilization between the self-managed and nurse-managed systems, because any impact through waiting times would be shared by all patients in the ED. Thus, the lack of effect on quality, revenue, and utilization between the two systems is more consistent with pure foot-dragging than the other mechanisms.

Any statement on quality, however, is limited by relatively imprecise estimates for mortality and bounce-backs. More direct evidence of foot-dragging is shown in the next section. Nevertheless, estimates for RVUs and costs are quite precisely estimated and rule out 2-5% relative changes. For example, given the current dollar conversion of about \$34 per RVU, the average ED patient represents about \$92 in revenue, while the effect of self-management on revenue is only -\$0.51 (95% CI -\$2.38 to \$1.36). With no change in revenue or costs per patient, delaying the discharge of patients thus unambiguously decreases productivity from a financial perspective.

Table 1.4.2 also reports the effect on the time to the first physician order, which I use as a proxy for the time between patient arrival at the pod and the patient first being seen by a physician. Significant free-riding would imply a positive coefficient for the self-managed system with respect to this proxy. However, the effect of the self-managed system on this measure is insignificantly different from 0 and slightly negative, indicating no significant free-riding.

The effect on length of stay due to Bravo's regime change to a self-managed system can also be seen graphically. Figure 1.4.1 shows month-year-pod fixed effects over time for the two pods estimated by this equation:

$$Y_{ijkpt} = \sum_{m=1}^M \sum_{y=1}^Y \alpha_{myp} I_{t \in m} I_{t \in y} + \beta X_{it} + \tilde{\eta}_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.4.2)$$

where the parameters of interest α_{myp} are fixed effects for each month, year, and pod interaction; $I_{t \in m}$ and $I_{t \in y}$ are indicator functions for t belonging to month m and year y , respectively; and $\tilde{\eta}_t$ is a revised sum of time category fixed effects that only includes day of the week and hour of the day. In Figure 1.4.1, there is a persistent discontinuity at the regime change that is consistent with my baseline estimates in Table 1.4.1 that self-managed teams in Bravo decreased patient lengths of stay. In addition, Figure 1.4.1 shows that trends in log length of stay in the two pods are conditionally parallel, which is the identifying assumption for Equation (1.4.1).

An issue that arises in difference-in-differences estimation is the construction of appropriate standard errors for inference (Bertrand et al., 2004).²⁷ My baseline specification clusters standard errors by physician, which is equivalent to an experiment sampling at the level of physicians, who are

²⁷This issue is largely only relevant to standard errors for the overall effect. Specific mechanisms use additional variation. In particular, foot-dragging relies on the effect of due to exogenous variation in expected future work.

given shifts that translate to pods and organizational systems, before and after the regime change in Bravo. This is the thought experiment I wish to consider, as the same physicians are in both pods before and after the regime change, and as shifts are assigned randomly, conditional on rough time categories.

Still, there is the additional statistical issue of unobserved and potentially correlated pod-level shocks over time. Therefore, I consider two alternative thought experiments for inference, both of which exploit the long time series and can be understood Figure 1.4.1. First, I address sampling variation at the pod level across time but with a more parametric form, assuming a month-year-pod shock that is correlated by a first-order autoregressive process across months within pod. Second, in the spirit of systematic placebo tests (Abadie et al., 2011; Abadie, 2010) and randomization inference (Rosenbaum, 2002), I consider the thought experiment that, under the sharp null of no effect of self-managed teams, there should be no significant difference between my obtained estimates and those I would obtain if I consider placebo regime changes over pod and month. This placebo approach considers randomization at the level of treatment rather than sampling. Detailed in Section 1.A.2, both approaches yield a high degree of statistical significance, with p -values less than 0.01.

1.5 Main Evidence of Foot-dragging

This section identifies the mechanism of foot-dragging with the following intuition: The expected gains to physicians by foot-dragging depend on expectations of future work. If no further patients arrive at the ED, then foot-dragging is not needed to prevent new work. But if there is an endless supply of patients waiting to be seen, then discharging a patient directly leads to having to see another one, and the incentive to foot-drag is extremely strong. I thus identify and quantify foot-dragging by increases in lengths of stay as expected future work increases. I measure expectations of future work as a function of the number of patients arriving at the ED.

Conceptually, other than through the moral hazard of foot-dragging, there is no other reason why lengths of stay should increase with expected future work, holding actual work constant.²⁸ I therefore interpret increasing length of stay with expected future work as evidence of foot-dragging. I hold this interpretation regardless of organizational system, but I am also interested in comparing foot-dragging between organizational systems. Also, as outlined in Section 1.3 above, it is important to note that the costs of foot-dragging derive from both direct moral hazard and inefficient assignment that results from this moral hazard. Both of these are directly related to expectations of future work, and both are jointly determined in equilibrium. I do not separately identify these effects here, but I separately address assignment in Section 1.7.

²⁸When both pods are open, patients that arrive at the ED while the index patient has just arrived at the pod may be sent to either pod. This allows me to separate expectations of future work (the number of patients arriving at the ED) from actual future work (the number of patients who will arrive at the pod). I discuss this further below.

For my baseline estimation of foot-dragging, I use equations of the following form:

$$Y_{ijkpt} = \alpha_1 EDWork_t + \alpha_2 Self_{pt} \cdot EDWork_t + \alpha_3 Self_{pt} + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.5.1)$$

for log length of stay Y_{ijkpt} for patient i , physician j , resident and nurse k , pod p , and time t . As before, $Self_{pt}$ indicates whether pod p at time t was self-managed, X_{it} controls for patient characteristics, ζ_p controls for time-invariant pod unobservables, η_t is a sum of time-category fixed effects (month-year interaction, day of the week, hour of the day), and ν_{jk} controls for the provider-trio.

$EDWork_t$ represents expected future work, defined in two ways. First, I consider ED arrival volume, defined as the number of patients arriving at triage in the hour prior to patient i 's arrival at the pod. The arrival of these patients is not controlled by physicians. They are seen by physicians via the computer interface, but their ultimate destination is unknown.²⁹ Second, I consider waiting room census, defined as the number of patients (the census) in the waiting room at the time of patient i 's arrival at the pod. Although physicians presumably can affect the number of patients waiting, this is a more salient measure of expected future work since physicians can readily click on the computer interface to see this census. The coefficients of interest in (1.5.1) are α_1 , α_2 , and α_3 . A positive α_1 indicates that physicians increase lengths of stay as expected future work increases (i.e., they foot-drag) in the nurse-managed system, while a negative α_2 indicates that the self-managed system mitigates foot-dragging. Coefficient α_3 represents the effect of the self-managed system after controlling for foot-dragging.

Table 1.5.1 reports estimates for (1.5.1), using both measures of expected future work – ED arrival volume and waiting room census. With each additional patient arriving hourly at triage or waiting in triage, lengths of stay increase by 0.6 percentage points in the nurse-managed system. The estimate of foot-dragging in the nurse-managed system is equivalent to a length-of-stay elasticity of 0.10 with respect to expected future work.³⁰ The coefficient for the interaction between expected future work and the self-managed system suggests that this effect is entirely mitigated in the self-managed system. That is, an additional patient in either measure of expected future work does not affect lengths of stay in the self-managed system. After controlling for foot-dragging, the coefficient representing the effect of the self-managed system is statistically insignificant in all specifications and ranges from -1% to 3%. In addition to the baseline specification in (1.5.1), I also include a number of controls for physician workload with pod-level volume at the time of patient arrival. Results are robust to including these controls.

²⁹Given average waiting times and the average length of stay, most patients will not even be assigned until after patient i is discharged. My results are also robust to alternative time windows for this measure.

³⁰I estimate this by using log measures of expected future work as $EDWork_t$. My preferred specification, shown in Table 1.5.1, does not take logs of expected future work because it is roughly normally distributed. However, results are qualitatively the same in this specification.

These results suggest substantial foot-dragging in nurse-managed teams and equally large mitigation of it in self-managed teams. Estimates are robust under both measures of expected future work – ED arrival volume and waiting room census. Controlling for actual current or future pod-level work does not change results, suggesting that increased patient lengths of stay are due to *expectations* of future work and insensitive to actual current or future workloads. Again, this distinguishes foot-dragging from other mechanisms of free-riding or advantageous selection. Without the moral hazard of seeking to prevent future work, physicians should not increase lengths of stay as expected future work increases. In contrast, the other mechanisms should only depend on actual work that reaches the pod.³¹ Finally, the remaining effect of the self-managed system is insignificant, suggesting that foot-dragging is quantitatively large enough to explain the difference in performance between self-managed and nurse-managed systems.

Figures 1.5.1 and 1.5.2 plot log-length-of-stay coefficients for each decile of expected future work interacted with organizational system, estimated by

$$Y_{ijkpt} = \sum_{d=2}^{10} \alpha_0^d (1 - Self_{pt}) D_d(EDWork_t) + \sum_{d=1}^{10} \alpha_1^d Self_{pt} D_d(EDWork_t) + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.5.2)$$

where $D_d(EDWork_t)$ equals 1 if $EDWork_t$ is the d^{th} decile, measured as ED arrival volume and waiting room census, respectively. The coefficients $\{\alpha_0^2, \dots, \alpha_0^{10}; \alpha_1^1, \dots, \alpha_1^{10}\}$ can be interpreted as the relative expected length of stay for patients in different organizational systems and under different states of expected future work, where the expected length of stay for patients in the nurse-managed system and under the first decile of expected future work is normalized to 0. As shown in Figures 1.5.1 and 1.5.2, lengths of stay progressively increase in the nurse-managed system as expected future work increases, which is consistent with the intuition that the incentive to foot-drag continues to grow as expected future work increases. The self-managed team has roughly the same expected length of stay as the nurse-managed team at low patient volumes, but its expected length of stay does not change with patient volume.

My measures of expected future work are likely to be noisy representations of physicians' true expectations of future work (e.g., they may expect more patients even when there is no one in the waiting room).³² Therefore this estimate is a lower bound on true foot-dragging: It is biased

³¹ Actual work is highly pod-specific. Two potential exceptions of spillovers between pods are waiting for a radiology test or a hospital bed. However, I find no difference in foot-dragging between patients based on likelihood, estimated by patient characteristics, to receive radiology tests. Table 1.5.2 also shows that radiology testing is unaffected by expected future work. The time spent waiting for a hospital bed is excluded from my measure of length of stay, since I record the time of the discharge *order*. In contrast, Table 1.5.2 shows that outcomes like admission, which are not supposed to be affected by foot-dragging, suggest spillovers from hospital congestion equally in *both* organizational systems.

³² Note also that I do not use expectations based on the usual volume for the time of the day or day of the week, which are absorbed by time fixed effects.

downward to the extent that I do not capture true expectations of future work. In addition, I interpret any increase in length of stay with expected future work as foot-dragging. But it may be reasonable to think that physicians in the absence of moral hazard should actually work *faster*, for example if they care about patients waiting too long in the waiting room. This is another sense in which my interpretation is a conservative benchmark: It assumes that, under no foot-dragging, there is either no attention to future work or no reason to work faster when future work increases. Note that since length of stay does not increase with expected future work in the self-managed system, foot-dragging relative to 0 and foot-dragging relative to self-managed teams are roughly the same in magnitude.³³

I also estimate the baseline equation for foot-dragging, Equation (1.5.1), for other outcomes and process measures: 30-day mortality, admissions, 14-day bounce-backs, RVUs, total direct costs, and a host of detailed process measures including laboratory, medication, and radiology orders. I show a subset of these outcomes and process measures in Table 1.5.2. I find no differential effect of expected future work between the two systems on any of these outcomes or process measures. Some outcomes do reflect a slight effect of ED arrival volume through hospital congestion for *both* systems, such as on hospital admissions and total costs, which include costs incurred in admissions. Foot-dragging effects on process measures are tightly estimated and show that the care provided while foot-dragging is not substantively different between the two systems. For example, under typical levels of ED arrival volume, I can rule out that the total order count increases by 0.4 in the nurse-managed system relative to the self-managed system, which represents a 3.3% increase against the average order count of 13. This is also consistent with pure foot-dragging, which delays the time of patient discharge but does not increase the quality or content of medical care.

1.6 Peer Effects on Foot-dragging

Foot-dragging could be reduced if physicians have more information about each other than the triage nurse has and if they also care about being seen foot-dragging. In this section, I explore peer effects on foot-dragging, as a joint test of more information between peers and social incentives, using three different types of analyses.³⁴

First, I examine foot-dragging when there is a peer in the pod compared to when there is no peer

³³Another support for this interpretation is the fact, shown in Section 1.6.1, that length of stay does not respond to expected future work when there is only one physician in the ED. With only one physician in the ED, there is no other physician present to foot-drag against. However, the physician may still foot-drag against future physicians, delaying discharge so that future work goes to physicians who are not yet there, and I technically cannot separate expected future work from actual future work, given that there is only one pod open.

³⁴Peer effects *on foot-dragging* are the effect of a peer interacted with expected future work; it isolates the effect of peers on foot-dragging moral hazard. In contrast, generic peer effects are simply the direct effect of a peer and could act through a variety of mechanisms, such as productivity spillovers. I discuss more generic peer effects in Section 1.A.3.1, where I show results similar to Mas and Moretti (2009) (i.e., productive peers increase the productivity of physicians).

in the pod but a physician working elsewhere in the ED. In the nurse-managed system, without social incentives, foot-dragging should not be affected by the location of other physicians. I use two additional settings – physicians working in a pod that is officially “self-managed” but without a peer, and physicians working alone in the entire ED – as falsification tests for my identification of foot-dragging. Second, I explore whether the type of peer present matters for foot-dragging in nurse-managed and self-managed teams as a test of heterogeneous social incentives across peer types. Third, I examine whether foot-dragging differs depending on a physician’s census and that of his peer, as a test of either strategic behavior or conditional social incentives.

1.6.1 Presence of a Peer

While physicians usually work in pods with a peer, during certain times on shift, physicians find themselves working without a peer. This is because shifts in each pod are staggered and because staffing adjusts upward in the morning when patient volume increases and downward in the night when it decreases. Such scheduling allows me to identify how physicians respond to the presence or absence of a peer.

In the nurse-managed system, only social incentives and more information between peers can explain the dependence of foot-dragging on the location of coworkers. That is, the total amount of future work and the number of physicians among whom to divide the work are sufficient statistics for foot-dragging unless if physicians care about being seen foot-dragging *and* if more information can be observed by peers working together in the same pod. In addition, two similar analyses can serve as falsification tests for the identification of foot-dragging by increases in length of stay with expected future work. First, when a physician in a self-managed pod is without a peer, he is effectively in a nurse-managed system: Every patient who arrives is in fact assigned to him by the triage nurse. The physician should then exhibit foot-dragging behavior as if working in the nurse-managed system (and without a peer). Second, when there is only one physician in the entire ED, there is essentially no assignment problem. That physician is responsible for all patients who arrive at the ED. With no coworker to foot-drag against, physicians have no incentive for foot-dragging.³⁵

In Table 1.6.1, I present results for regressions of the form

$$Y_{ijkpt} = \alpha_1 EDWork_t + \alpha_2 NoPeer_{jt} \cdot EDWork_t + \alpha_3 NoPeer_{jt} + \beta X_{it} + \zeta_p + \eta_p + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.6.1)$$

for nurse-managed-team and self-managed-team samples separately. $EDWork_t$ is ED arrival volume, or the number of patients arriving at ED triage in the hour prior to the index patient’s arrival at

³⁵Physicians can still foot-drag against future physicians, but this theoretically is no different at any other time. I cannot control for actual future work in this scenario, because all work that comes to the ED eventually goes to the same pod and physician. However, this would only bias measured foot-dragging upwards if the omitted volume of actual work is positively correlated with expected future work and increases length of stay.

the pod, and $NoPeer_{jt}$ is a dummy for whether physician j has no peer in the same pod. These regressions test behavioral responses by physicians within a system, depending on peer presence. I estimate Equation (1.6.1) only when there are at least two physicians in the ED, so that foot-dragging always entails a negative externality against a current coworker. I also perform the pooled regression

$$Y_{ijkpt} = \sum_{s=1}^4 1(PeerState_{jt} = s)(\alpha_s EDWork_t + \delta_s) + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.6.2)$$

which estimates the degree of foot-dragging with coefficient α_s for each of four peer states $s \in \{1, \dots, 4\}$: alone in a pod but not alone in the ED, with a peer in the nurse-managed system, with a peer in the self-managed system, and alone in the ED.

Results in Table 1.6.1 are consistent with previously estimated coefficients for the increase in length of stay with expected future work, shown in Table 1.5.1, in both systems when a peer is present. When a peer is not present, however, length of stay increases much more quickly with expected future work. Estimates in the first two columns suggest that, without a peer present, the response to expected future work quintuples in the nurse-managed pod and increases in the self-managed pod (but effective nurse-managed system) to almost triple the magnitude as in the nurse-managed system with a peer. Results from the pooled regression in Equation (1.6.2), shown in the third column of Table 1.6.1, confirm this and show that physicians do not increase lengths of stay with expected future work when they are alone in the ED.

These results suggest that physicians reduce their foot-dragging when a peer is present, which is consistent with social incentives and the fact that peers can observe each other's true workloads better than physicians in different locations. Additionally, results from the two falsification tests, in the second and third columns of Table 1.6.1, are consistent with the interpretation of increases in length of stay with expected future work as foot-dragging, a moral hazard reduced by social incentives and self-management. First, physicians working without a peer in an officially self-managed pod (but effectively nurse-managed system) engage in foot-dragging to triple the extent of those working with a peer in a nurse-managed pod. Second, when a single physician is responsible for all patients entering the ED, I find no evidence of foot-dragging.

It is unlikely that these large effects can be explained by anything other than peer effects on foot-dragging. First, both regressions (1.6.1) and (1.6.2) include time dummies for hour of the day. More generally, the coefficients of interest in these regressions (and all other regressions identifying foot-dragging) correspond to responses of length of stay to expected future work, rather than levels of length of stay. Second, essentially all observations with only one physician in a pod occur during transition times of two to three hours during shifts in which the same physician works with a peer. The effect is thus identified by behavior of the same physician in the same shift, but only under different peer conditions. Third, ED conditions are generally unchanged in this short window of

time relative to nearby times. One obvious change when a peer leaves is that the ED workload is distributed among physicians who are one fewer. However, results are qualitatively unchanged when normalizing measures of expected future work for the number of physicians in the ED.

1.6.2 Peer Type

Given that foot-dragging is reduced by the presence of a peer, I next consider different types of social relationships between physicians and their peers. Since social connectedness, driven by demographics or shared history, has been shown to influence peer effects (Bandiera et al., 2010, 2005; Jackson and Schneider, 2010), I first consider peers of the same sex, similar age, or same place of residency training as potentially more connected to each other. Second, I consider peers who are faster (or more productive) than median. The effect of this peer type on foot-dragging may include both social and strategic concerns. To see the strategic concerns, note that slower peers will cause more work to be redirected to physicians unless they slow down as well.³⁶ Third, I consider peers, by their history of time working with each other, who are more familiar with each other's workplace behavior and more likely to have established reputations with each other.³⁷ Finally, I consider peers who have at least two years greater tenure than the index physician. Social hierarchy is a common feature in many workplaces, particularly those with professionals, long tenures, or strong work cultures. To my knowledge, the peer effects due to hierarchical social relationships has not yet been studied in economics.

For each of these peer types, I estimate regressions of the following form:

$$Y_{ijkpt} = \alpha_1 EDWork_t + \alpha_2 PeerType_{jt}^m \cdot EDWork_t + \alpha_3 PeerType_{jt}^m + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.6.3)$$

separately for nurse-managed and self-managed samples. $PeerType_{jt}^m$ is an indicator that the peer for physician j at time t is of type m . I am interested in the coefficient α_2 as the effect of working with a peer type on foot-dragging, again identified by increases in length of stay with respect to expected future work.

Table 1.6.2 reports results for three of the peer types.³⁸ Senior peers are the only peer type showing a significant effect on foot-dragging. In the nurse-managed system, working with a senior peer decreases foot-dragging by half, from an increase of 0.8% for each patient arriving at the ED to an increase of 0.4%. In a pooled regression shown in the third column, it also appears that

³⁶In addition to these mechanisms for peer effects on foot-dragging, there are other mechanisms for direct peer effects. See footnote 34 and Section 1.A.3.1.

³⁷I use a threshold of at least 60 hours working in the same pod, which is at the 75th percentile, to describe peers that are “familiar” with each other. Given physician turnover and a large number of shift times and locations, it is relatively uncommon for two ED physicians to have longer histories working together in the same pod.

³⁸For brevity, I omit peer types related to social connectedness from Table 1.6.2, as they show no effect on foot-dragging.

senior peers further reduce foot-dragging in the self-managed system.³⁹ Other peer types – highly productive peers, familiar peers, and connected peers – show no significantly differential effect on foot-dragging. These results suggest that the most important social relationships between peers may be unilateral ones based on hierarchy, as opposed to ones that are based on connectedness or familiarity.

The effect of peer relationships on efficiency may have implications for the construction of teams. Of course, this depends on what relationships are most important. If hierarchical relationships are most important in mitigating foot-dragging, then one possible implementation will be to have teams composed of physicians with different tenures. However, this effect appears small relative to the effect of having any peer present or of the self-managed system.

1.6.3 Physician and Peer Censuses

As physicians in the nurse-managed system both foot-drag and pay attention to the presence and identity of their peers, it follows that foot-dragging may also depend on the relative amount of work peers have (and beliefs about whether they might be foot-dragging). As discussed above, an important summary statistic for the state of work in a pod is the distribution of patients between a physician and his peer.

Foot-dragging as a policy that depends on own and peer censuses could be influenced by both strategic behavior and conditional social incentives. Strategically, in the nurse-managed system, censuses influence patient assignment by the triage nurse. For example, in the theoretical framework, physicians with high censuses should have less of an incentive to foot-drag because they are already unlikely to receive patients. On the other hand, socially, foot-dragging may be less acceptable to peers at certain joint censuses. In particular, an experimental literature has convincingly shown “reciprocity” in games with public goods (Fehr and Gächter, 2000), which can be explained by a social aversion toward inequity (Fehr and Schmidt, 1999).

For now, I aim to test for either strategic behavior or social incentives that are conditional on the distribution of work by measuring foot-dragging conditional on physician and peer censuses. I consider $Census_{jt}$, defined as the number of patients being cared for by physician j at time t , and the corresponding $Census_{-jt}$ for his peer $-j$.⁴⁰ I summarize these censuses into quintiles, with QU_{jt}^m denoting that physician j ’s census at time t is in the m^{th} quintile, and estimate the following regression:

³⁹The pooled regression includes interactions with the self-managed system and a direct effect for the self-managed system. I do not write out this equation above, as Equation (1.6.3) communicates the effect of interest in α_2 . Recall that self-management and social incentives may independently reduce foot-dragging, and that there may be some foot-dragging in the nurse-managed system, since the 0 benchmark for no foot-dragging is conservative.

⁴⁰I assume that patients appear on physician j ’s census once they arrive at the pod. This is certainly true in the nurse-managed system. In the self-managed system, it abstracts away from the fact that physician j has to choose patients on his census when working in self-managed teams.

$$Y_{ijkpt} = \sum_{m=1}^5 \sum_{n=1}^5 QU_{jt}^m QU_{-jt}^n (\alpha_1^{m,n} EDWork_t + \alpha_2^{m,n}) + \beta X_{it} + \zeta_p + \eta_t + \nu_{vk} + \varepsilon_{ijkpt}. \quad (1.6.4)$$

$EDWork_t$ is again measured as ED arrival volume and represents expected future work. The general coefficient of interest, $\alpha_1^{m,n}$, represents the degree of foot-dragging by physician j when his census is in the m^{th} quintile and his peer's census is in the n^{th} quintile. As before, I interpret foot-dragging as increases in length of stay with expected future work. I estimate this model separately for nurse-managed and self-managed teams; I am interested in testing for heterogeneity in foot-dragging in the nurse-managed system, while I am mostly using the self-managed system as a falsification test for heterogeneity, as I do not expect and have not found significant foot-dragging in the self-managed system.

Table 1.6.3 presents estimated foot-dragging coefficients $\alpha_1^{m,n}$ from (1.6.4). These estimates reveal several features in the two organizational systems. First, using 0 as the benchmark for no foot-dragging, there is virtually no foot-dragging in the self-managed system, regardless of censuses. This is consistent with previously discussed findings of little to no foot-dragging in the self-managed system. Second, foot-dragging in the nurse-managed system shows remarkable dependency and predictability according to the exact joint-census state (also shown in Appendix Figure 1.A.5). As predicted by strategic behavior, physicians foot-drag more when they have lower censuses. Although I do not separate strategic behavior from social incentives in this reduced-form analysis, physicians notably refrain from foot-dragging when both censuses are “normal” (i.e., in the third quintile), which suggests conditional cooperation. On the whole, however, physicians fail to refrain from foot-dragging in the nurse-managed system.

1.7 Patient Assignment and Equilibrium Building

I have shown that the self-managed system improves throughput productivity by reducing foot-dragging and that physician peers in the same pod observe more information about each other's true workloads. In this section, I test a prediction related to the *use* of this information by physicians in the self-managed system to assign patients.

According to the management literature, self-managed teams improve efficiency by both “monitoring and managing work” (Pallak and Perloff, 1986). As formalized in the model discussed in Section 1.3, if physicians use more information about true relative workloads to choose patients in the self-managed system, foot-dragging should not only decrease, but the *ex post* assignment efficiency should also increase. That is, patients should be assigned *more* often in the self-managed system than in the nurse-managed system according to censuses as signals of workload.

In addition to studying patient assignment in steady state, a related but distinct issue is how

physicians built the new equilibrium after Bravo pod switched from a nurse-managed to a self-managed system.⁴¹ In Figure 1.7.1, I show that foot-dragging does not immediately disappear after the switch in Bravo, by estimating the effect of expected future work, interacted with four-month interval dummies, on length of stay. Rather, it takes at least five months to disappear. Therefore, a second question relates to the assignment of patients during this transition period, and in particular whether foot-dragging physicians are more likely to get new patients.⁴²

To address these questions, I study the relationship between censuses and patient assignment. Censuses are a public measure of workload, but they are signals that can be distorted by foot-dragging. A testable implication of the prediction of *ex post* assignment inefficiency is that physician censuses should be *less* related to new patient assignment in the nurse-managed system than in the self-managed system in equilibrium. Stated another way, this tests the theory that signals are used more efficiently in the self-managed system. Assignment during the off-equilibrium transition of Bravo pod sheds light on how the equilibrium with no foot-dragging is eventually achieved in the self-managed system. Specifically, patient assignment to physicians who foot-drag should result in a positive correlation between assignment and censuses.

I measure the correlation between censuses and new patient assignment for both pods over time. For my baseline specification, I estimate the linear probability model

$$I_{ijt} = \alpha \text{Census}_{jt} + \beta \text{Shift}_{jt} + \eta_j + \nu_{it} + \varepsilon_{ijt}, \quad (1.7.1)$$

where the outcome I_{ijt} is an indicator variable for whether patient i , who arrives on the pod at time t , is assigned to physician j . Census_{jt} denotes the number of patients under the care of physician j at time t and is the variable of interest. Shift_{jt} includes time indicators of physician j 's shift at time t , since physicians are less likely to be assigned new patients as they near the end of their shift, regardless of their censuses. The fixed effect η_j controls for physician identities and allows for some physicians being more likely to take new patients regardless of census or observed behavior by their peers. The fixed effect ν_{it} , controlling for patient i at visit t , ensures that two physicians could have been assigned the patient. It implies that this linear probability model is equivalent to estimating a differenced model in which the variable of interest is $\text{Census}_{jt} - \text{Census}_{-jt}$, the difference in censuses between a physician and his peer, with a coefficient algebraically equal to α . The coefficient α represents the incremental likelihood, averaged over different shift times, with which a physician is to receive a new patient for each additional patient on his census relative to

⁴¹As discussed in Gibbons and Henderson (2012), equilibria may not be immediate or obvious, and they instead might need to be “built.” They categorize reasons for this under needs to establish “credibility” and “clarity,” and they review theoretical, experimental, and case-study literature on this (e.g., Greif, 1993; Fudenberg and Levine, 1993; Weber and Camerer, 2003). I discuss this issue of why the equilibrium in Bravo after its regime change is not immediate more specifically later in this section.

⁴²For more discussion of how patients are assigned in the self-managed system, see Section 1.3.3 and Section 1.B.3. Regardless of whether physicians can commit to an assignment policy and of the amount of information they observe, patients should be assigned to physicians who foot-drag with greater probability if they have lower true workloads.

his peer's census.

Figure 1.7.2 shows a plot of coefficients α in (1.7.1) over time and in both pods.⁴³ I estimate α over each month by using triangular kernels with 45 days on each side of the first of the month; for months immediately before and after the regime change, I only use 45 days on the side away from the regime change. Prior to the Bravo regime change, Figure 1.7.2 shows relatively stable assignment in both pods. In both systems, physicians with lower censuses are more likely to be assigned patients, but this likelihood is consistently greater in the self-managed system (Alpha) than in the nurse-managed system (Bravo). Immediately after the regime change in March 2010, assignment in Bravo shows a jump in which physicians with *higher* censuses are actually more likely to receive new patients. After three months, the spike reverses, and patients are again more likely to be assigned to physicians with lower censuses, even more so than prior to the regime change. Finally, Bravo's correlation between assignment and censuses after the spike settles to the same level as Alpha's. The qualitative features of these assignment policy functions are robust to a number of different specifications, including logit estimation, omission of physician fixed effects η_j , alternative kernel bandwidths, and controlling for other workload observables such as average patient severity.

These results show that, in equilibrium, the self-managed system improves the *ex post* assignment efficiency according to publicly observable signals of workload. By measuring assignment in both pods over time, I also show that assignment is not specific to pods, but rather to the organizational systems. This is consistent with the theory in Section 1.3: As the threat of foot-dragging is reduced in the self-managed system by the use of more information between peers, new patients can be more readily assigned to physicians with lower censuses.⁴⁴ Stated differently, foot-dragging is not reduced in the self-managed system by a trivial mechanism that ignores relative workloads.

Answering the second question – of how the new equilibrium is built during the transition period after the Bravo regime change – is admittedly more speculative. Most standard models, including mine, are silent on building new equilibria, since behavior is considered already *in equilibrium*. Nonetheless, a well-known experimental regularity is that players do not usually settle immediately on equilibria. Experiments have shown that people are influenced by the mere name of a game (Liberman et al., 2004) or irrelevant past experiences that could lead to “institutional afterglow” (Bohnet and Huck, 2004). Specific to this study, although physicians had worked in self-managed teams in Alpha, the regime change in Bravo was not announced as a move to replicate the organizational system in Alpha, but rather as a simple merger of bed ownership between peers with nothing else changed (e.g., Bravo was still closed at night). Additionally, as mentioned in Section 1.2, most ED physicians work sporadically, and so some would have been new to the regime change in Bravo even after a few weeks after implementation. Finally, the self-managed system requires physicians to work more as a team, and therefore outcomes are more dependent on the beliefs and strategies

⁴³See Appendix Figures A-4.2 and A-4.3 for a set of plots representing the same estimates with confidence intervals.

⁴⁴Although many triage nurses are not aware that they are optimizing *ex ante* assignment by ignoring censuses under the threat of foot-dragging, they are given management “rules of thumb” to alternate assignment unless censuses are sufficiently out of balance.

of both peers.

Given that full cooperation is not immediate, the question is how full cooperation is eventually established. Again, some insight can be gained from the experimental literature. Enforcement in public goods games has been studied in seminal research by Ostrom et al. (1992) and Fehr and Gächter (2000). They have found that full cooperation is possible only when players are allowed to enforce it by punishment. In other words, the rules of the game, akin to the organizational environment, determine the degree of cooperation even with the same players and same information. Unlike laboratory experiments, this study cannot definitively show punishment. I do however see that, during the transitional period of residual foot-dragging in Bravo, physicians with higher censuses were *more* likely to be assigned new patients, reflected in the spike in Figure 1.7.2. This is consistent with foot-dragging physicians being made to take new patients, possibly because their foot-dragging was observed by peers who decided not to take new patients upon observing that they already bore higher true workloads. Such an assignment policy is consistent with eventually building a new equilibrium with no foot-dragging.

1.8 Conclusion

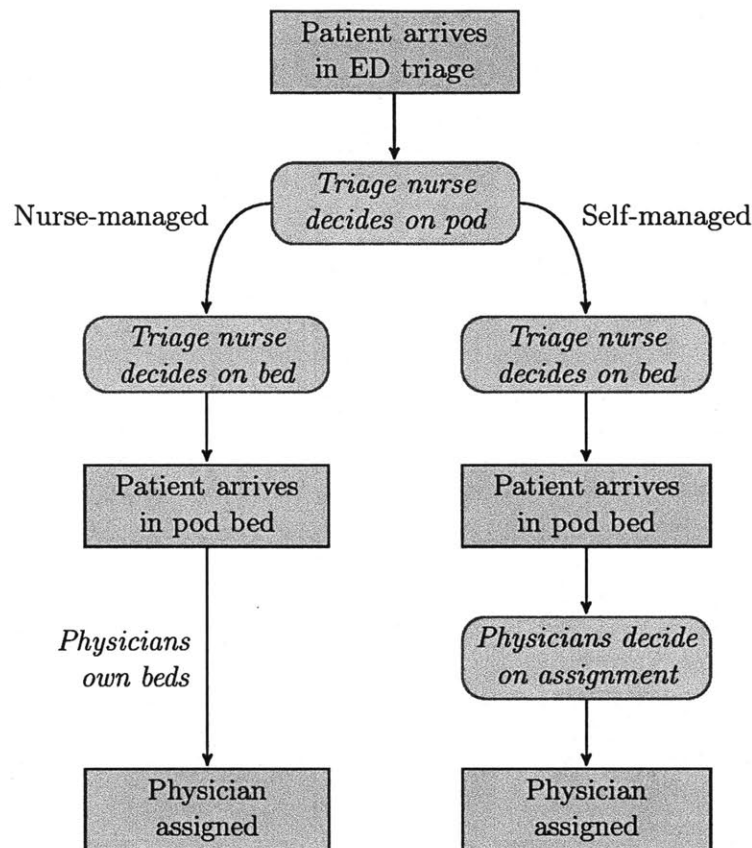
This paper sheds light on how organizational structure improves physician worker productivity. In particular, this study draws a contrast between two systems in which physicians have different managerial authority: In the nurse-managed system, physicians are assigned work by a triage nurse “manager”, while in the other self-managed system they are responsible for assigning work among themselves. The self-managed system produces significantly shorter patient lengths of stay than the nurse-managed system, primarily via the reduction of a moral hazard I call foot-dragging, in which physicians delay patient discharge to forestall new work. Even in the nurse-managed system, the presence of peers reduces foot-dragging, suggesting both social incentives and more information between peers. The self-managed system appears to use this information between peers further to improve patient assignment and reduce foot-dragging.

Although I find that foot-dragging essentially explains the overall effect of self-managed systems, I do find preliminary evidence that physicians choose patients differentially and that there may be distributional effects across patient types, depending on desirability to physicians. This is beyond the scope of this paper and is an area of further research. For practical reasons, I focus on observable measures of productivity, but there may be a host of other considerations, such as conflict among physicians, that could arise from self-management but would be more difficult to measure in this study. However, many of these issues have been addressed extensively in literature outside of economics and point quite favorably toward self-management (Yeatts and Hyten, 1997; Pallak and Perloff, 1986).

This study examines a precise but relatively minor experiment in the arrangement of ED physician teams that followed from a simple, discrete change in the ownership of beds. In practice,

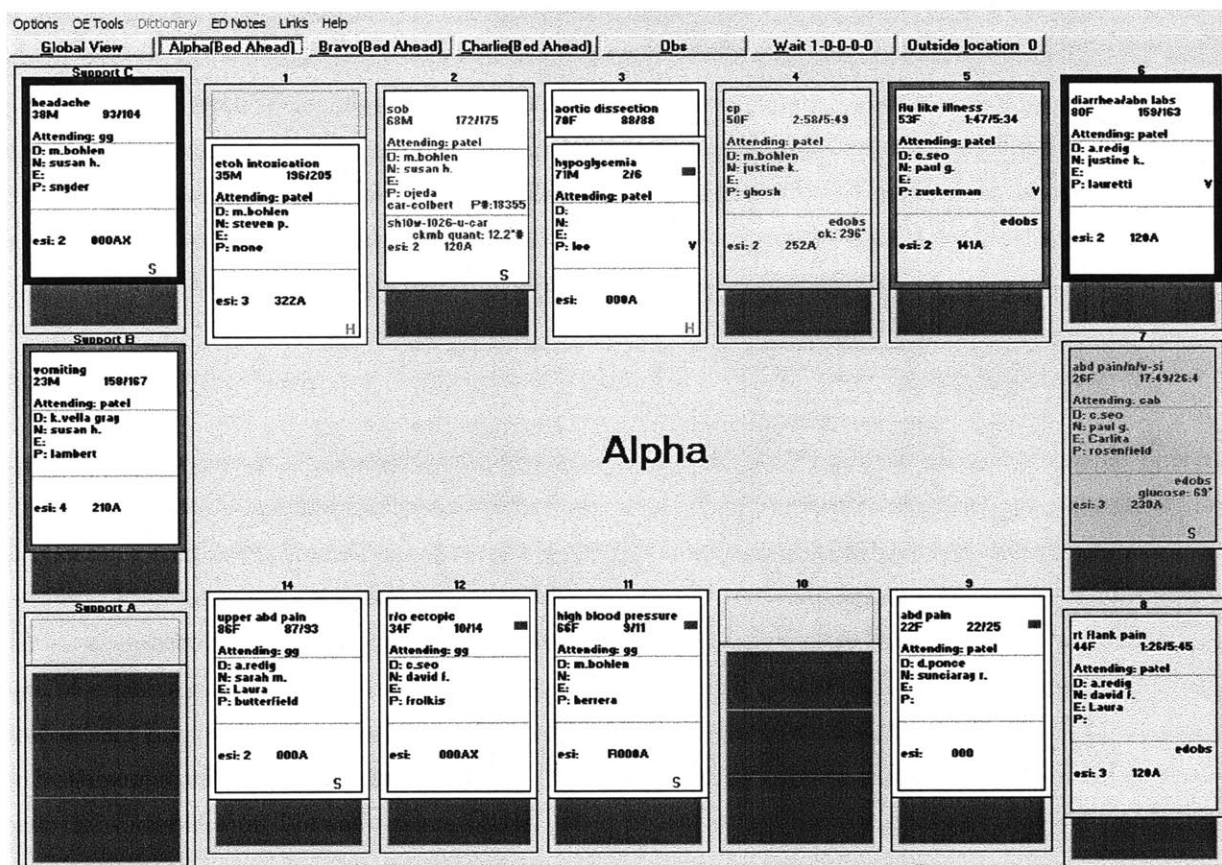
however, self-managed teams and organizations in general differ in a wide variety of ways, including for example the number of workers in the same team and the nature of work relationships among team members. Indeed, there is substantial variation in organizational structure even within health care. Other arrangements, particularly those with a more obvious focus on teamwork, may have significantly different effects on the behavior of workers and should be further studied.

Figure 1.2.1: Patient-to-physician Assignment Algorithm



Note: This figure shows the patient assignment algorithm, starting with patient arrival at ED triage and ending with assignment to a physician. In ED triage, the triage nurse decides which pod and bed to send the patient. If the triage nurse decides to send the patient to a pod with a nurse-managed system (if one exists), then she also makes the decision on which physician will be assigned the patient because physicians own beds. If she decides to send the patient to a pod with a self-managed system, then she does not assign the physician. The physicians currently working in the self-managed pod will decide among themselves on that assignment. Although not shown in the figure, the triage nurse always assigns the bed and the nurse; she never assigns the resident, since residents in either pod choose their own patients or are told by physicians to see patients.

Figure 1.2.2: Computer Schematic of Alpha Pod



Note: This figure shows a computer screen layout of Alpha pod, which is both a geographic representation of the physical pod and the interface for physicians to select patients, examine the electronic medical record, and enter orders. Slots represent beds, with two beds per room, and filled beds are represented by slots with information. Colors represent various patient states, for example, whether an order needs to be taken off or whether the patient has been ordered for discharge. Identifying patient information has been removed here, but when displayed, such information includes patient last name, chief complaint, age, sex, physician, resident, nurse, emergency severity index, and minutes in ED (including triage) and in pod. Alpha is always self-managed.

Figure 1.2.3: Computer Schematic of Bravo Pod

Options OE Tools Dictionary ED Notes Links Help

Global View Alpha(Bed Ahead) Bravo(Bed Ahead) Charlie(Bed Ahead) Obs War 0-0-0-0-0 Outside location 0

Support 26

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

K axilla lump
31F 51/02
Attending: brooks
D: p.montaz
N: patricia m.
E: P: modest
esi: 3 000A

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

gl bleeding
46M 61/69
Attending: brooks
D: n.kostarellas
N: robert b.
E: P: pollack
esi: 2 020A S

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

vomiting
55F 0:2014:2
Attending: bhatia
D: j.healy
N: robert b.
E: P: christopher
esi: 3 000A S

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

rtio dka
30F 5:34/5:34
Attending: bhatia
D: n.askman
N: robert b.
E: katie m.
P: morris-singer
glucose: 348*
512A
esi: 2 S

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

chest pain
74F 6:06:11
Attending: bhatia
D: p.montaz
N: stephanie m.
E: katie m.
P: coken
dexter-balley P#-30524
sh3w-914-u-dexter
esi: 2 212A

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

newid
61F 230/240
Attending: bhatia
D: n.askman
N: stephanie m.
E: P: ven
one-feldman P#-30719
own7n-730-c-one
esi: 2 R342A S

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

weakness
93F 17/21
Attending: bhatia
D: n.kostarellas
N: brianna v.
E: P: phull
esi: 000A

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

infected facial wound
36M 04/100
Attending: brooks
D: leasey
N: brianna v.
E: P: post
esi: 3 000A

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

bloody diarrhea
20F 0:4778:59
Attending: kayden
D: p.montaz
N: mary o.
E: NICK
P: none
med-sukul P#-31078
14c-521-c-med
esi: 3 130

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

abd pain
48M 102/101
Attending: bhatia
D: n.kostarellas
N: mary o.
E: NICK
P: mosveeney
esi: 3 R1152A

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

pneumonia
66M 7:19/7:19
Attending: kayden
D: leasey
N: lisa a.
E: NICK
P: buckle
med-sukul P#-31078
14c-201-u-med
esi: 2 152A

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

chest pain
43M 153/203
Attending: bhatia
D: n.askman
N: mary o.
E: NICK
P: satyam
esi: 4 000AX H

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

43

42

41

40

39

38

Support 27

Bravo

32

33

34

35

36

Support 37

Support 26

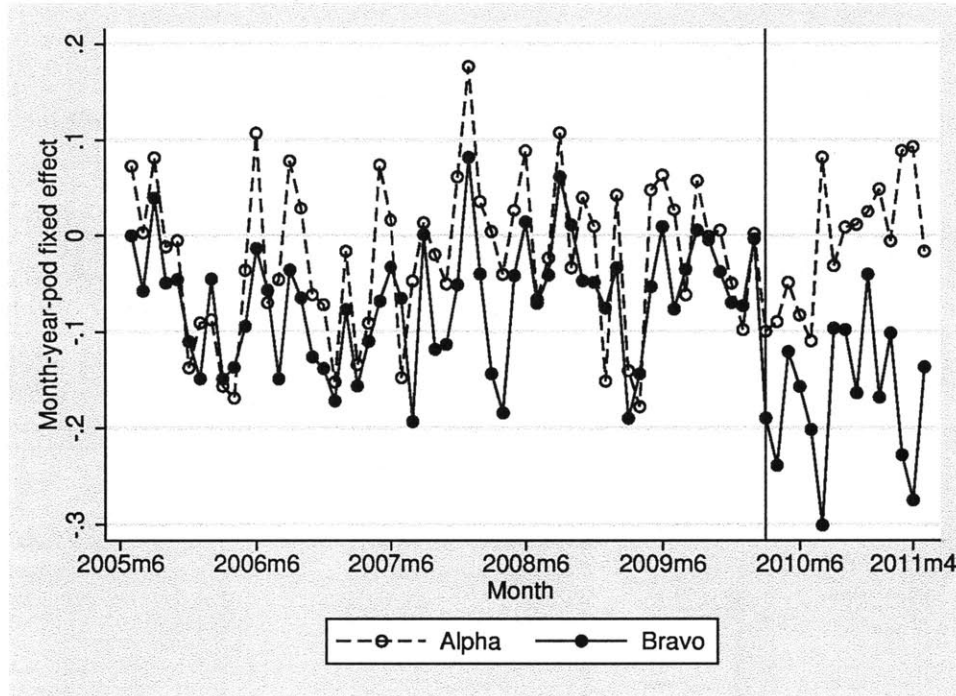
leg bruise/pain
76M 105/120
Attending: bhatia
D: n.askman
N: mary o.
E: NICK
P: satyam
esi: 4 000AX H

Support 27

inf neophrostomy tube
61M 161/159
Attending: bhatia
D: leasey
N: patricia m.
E: conale
P: prior
med-aaraso P#-30795
14c-561-o-med
vof
esi: 3 342A

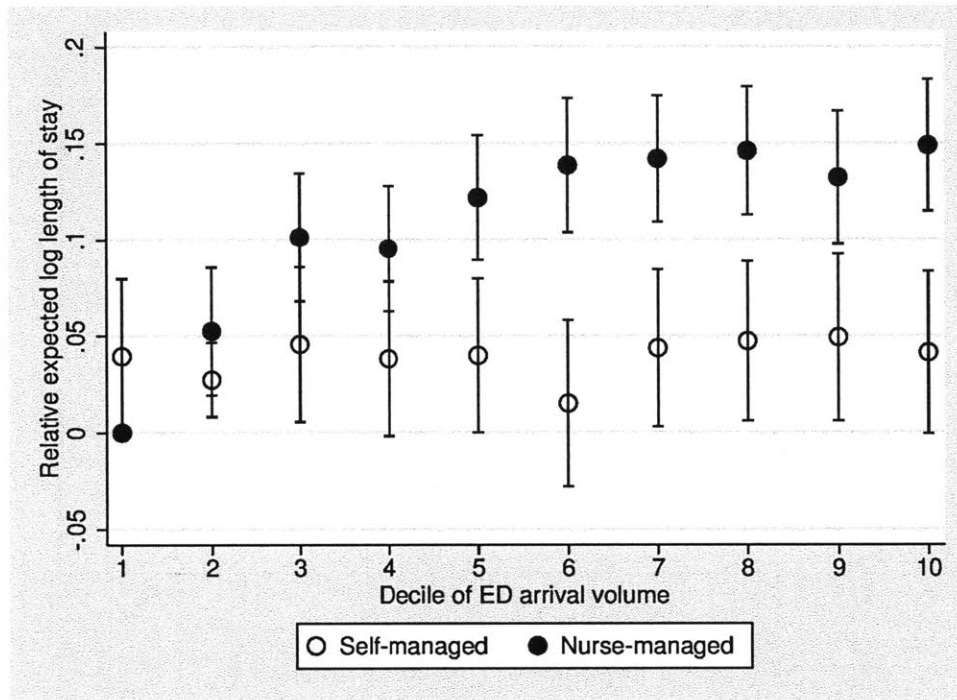
Note: This figure shows a computer screen layout of Bravo pod, which is both a geographic representation of the physical pod and the interface for physicians to select patients (when applicable under the self-managed system), examine the electronic medical record, and enter orders. Information represented is the same as for Alpha pod, shown and explained in Figure 1.2.2. Bravo was a nurse-managed system then changed to a self-managed system in March 2010. When Bravo pod was under the nurse-managed system, the geographic layout and the screen was simply divided in half, generally with one physician occupying each half. The image above was taken with Bravo under the self-managed system.

Figure 1.4.1: Overall Effect of Self-managed System on Log Length of Stay



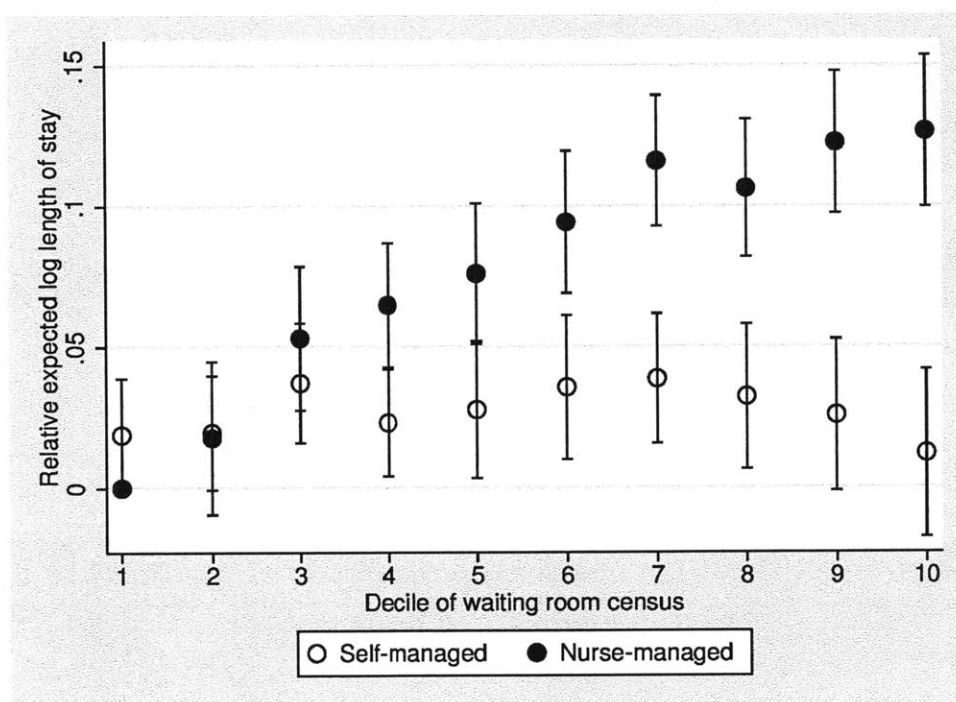
Note: This figure shows month-year-pod fixed effects estimated in a regression of log length of stay, as in Equation (1.4.2). Alpha pod fixed effects are plotted with hollow blue circles; Bravo pod fixed effects are plotted with solid red circles. The vertical red line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed. The fixed effect for Bravo in the first month is normalized to 0. The regression controls for uninteracted ED arrival volume, time categories (dummies for month-year, hour of the day, and day of the week), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions.

Figure 1.5.1: Foot-dragging as Expected Future Work (ED Arrival Volume) Increases by Deciles



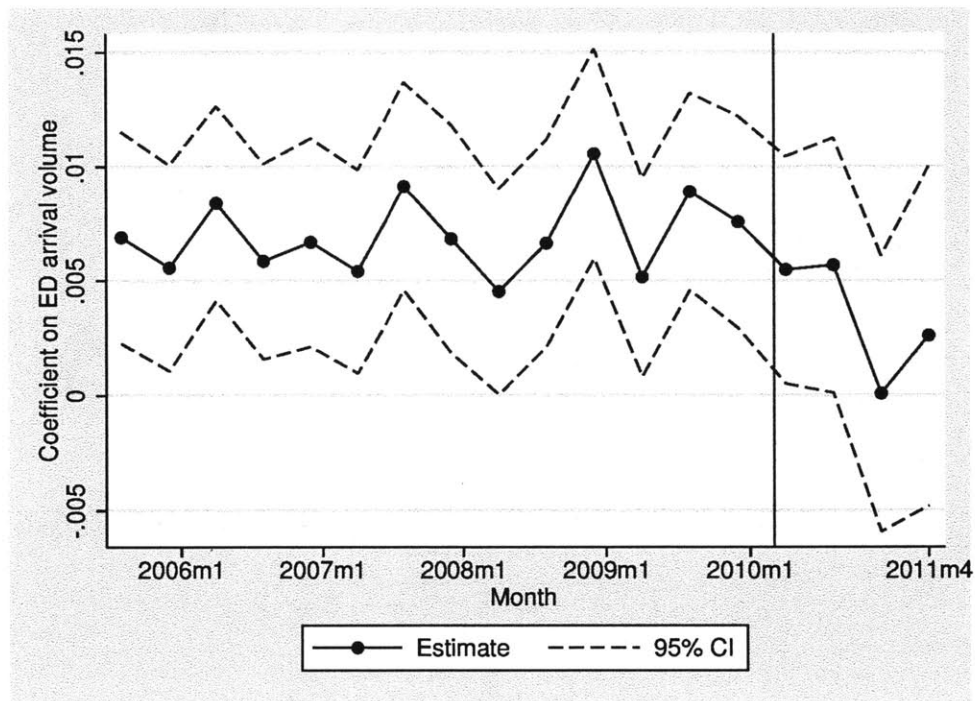
Note: This figure shows relative expected log length of stay as a function of expected future work, as measured in deciles of ED arrival volume, or the number of patients arriving at triage in the hour prior to the patient's arrival at the pod. Expected log length of stay is normalized to 0 in the nurse-managed system and with the first decile of ED arrival volume. Coefficients for these decile-pod dummies are plotted from the regression of (1.5.2). Hollow blue circles indicate coefficients for the nurse-managed system; solid red circles indicate coefficients for the self-managed system. Vertical brackets show 95% confidence intervals. All models control for time categories (dummies for month-year, hour of the day, and day of the week), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), and physician-resident-nurse interactions. Results are insensitive to controlling for pod-level patient volume.

Figure 1.5.2: Foot-dragging as Expected Future Work (Waiting Room Census) Increases by Deciles



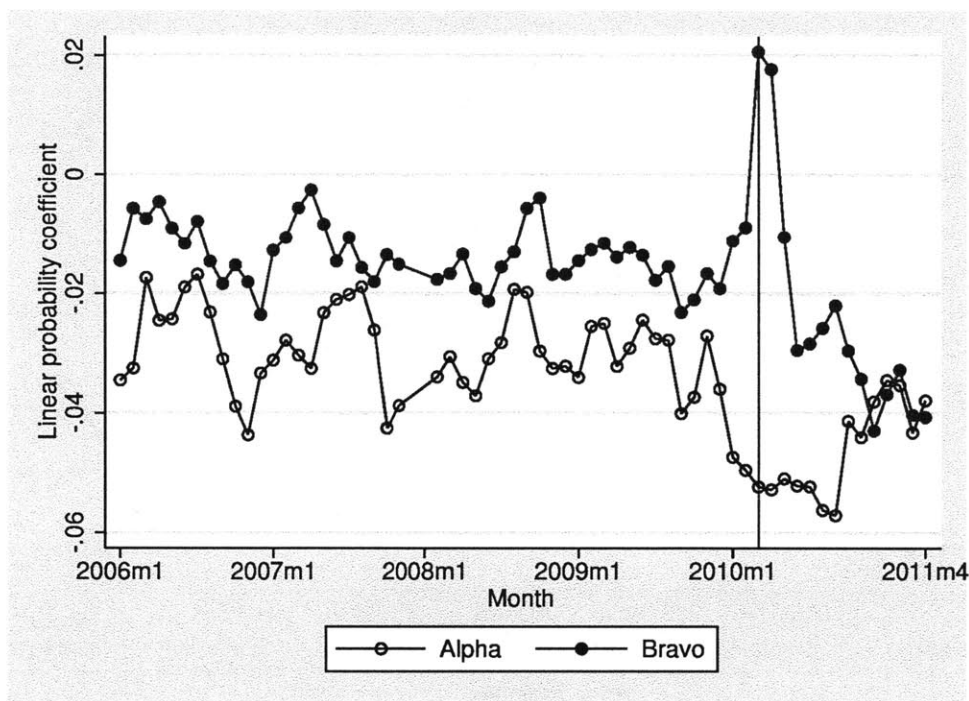
Note: This figure shows relative expected log length of stay as a function of expected future work, as measured in deciles of waiting room census at the time of the patient's arrival at the pod. Expected log length of stay is normalized to 0 in the nurse-managed system and with the first decile of waiting room census. Coefficients for these decile-pod dummies are plotted from the regression of (1.5.2). Hollow blue circles indicate coefficients for the nurse-managed system; solid red circles indicate coefficients for the self-managed system. Vertical brackets show 95% confidence intervals. All models control for time categories (dummies for month-year, hour of the day, and day of the week), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), and physician-resident-nurse interactions. Results are insensitive to controlling for pod-level patient volume.

Figure 1.7.1: Event Study of Foot-dragging in Bravo



Note: This figure shows foot-dragging in Bravo pod, as estimated by the log-length-of-stay coefficient for expected future work (measured by ED arrival volume, defined as the hourly rate of patients arriving at triage when the index patient arrives at the pod) interacted with four-month interval dummies. These coefficients are plotted as an event study before and after the regime change of Bravo pod from a nurse-managed system to a self-managed system in March 2010, shown with a vertical red line. The solid dots plot the coefficient estimates, and the dotted lines plot the 95% confidence intervals.

Figure 1.7.2: Effect of Additional Census Patient on New-patient Assignment Probability



Note: This figure shows the new-patient assignment probability, as a function of relative censuses for physicians within each pod. The plotted coefficient estimates represent the average effect on assignment probability of each additional patient on a physician's census relative to his peer's census. Hollow blue circles show coefficient estimates for Alpha pod, which was always self-managed. Solid red dots show the coefficient estimates for Bravo pod, which switched to a self-managed system in March 2010, shown with a vertical red line. Coefficients are estimated in a kernel regression using a triangular kernel with 45 days on each side; estimates for February and March 2010 in Bravo pod are estimated by a kernel with 45 days only on the same side of the regime change. For simplicity, 95% confidence intervals are not plotted; see Appendix Figures A-4.2 and A-4.3 for plots with confidence intervals.

Table 1.2.1: Patient Characteristics Available for Physicians with Above- or Below-median Productivity

Patient characteristic	Physicians with above-median productivity	Physicians with below-median productivity
Age	48.7 (19.6)	48.7 (19.6)
Emergency severity index	2.74 (0.78)	2.74 (0.78)
White	0.508 (0.500)	0.509 (0.500)
Black or African-American	0.233 (0.423)	0.234 (0.423)
Spanish speaking	0.098 (0.297)	0.097 (0.296)
Female and age < 35 years	0.187 (0.390)	0.185 (0.388)

Note: This table reports averages and standard deviations (in parentheses) for each characteristic of patients available to physicians with above-median and below-median productivity. “Available” means available to choose from in the self-managed system or assigned to in the nurse-managed system. Physician productivity is estimated by fixed effects in a regression of length of stay, controlling for all possible interactions of team members (physician assistant or resident and nurse), coworker, and pod; patient characteristics (age, sex, emergency severity index, Elixhauser comorbidity indices); ED arrival volume; and time categories (month-year, day of the week, and hour of the day dummies). The average difference in productivity between physicians of above- and below-median productivity is 0.28, meaning that physicians with above-average productivity have 28% shorter lengths of stay than those with below-average productivity.

Table 1.2.2: ED Conditions for Physicians with Above- or Below-median Productivity

Prior patient volume for entire ED	Physicians with above-median productivity	Physicians with below-median productivity
<i>During any shift</i>		
Within last hour	6.06 (3.87)	5.97 (3.86)
Within last 6 hours	34.90 (19.11)	35.15 (18.95)
<i>While in self-managed team</i>		
Within last hour	5.63 (3.77)	5.42 (3.73)
Within last 6 hours	34.46 (17.98)	34.33 (17.63)
<i>While in nurse-managed team</i>		
Within last hour	6.40 (3.92)	6.40 (3.90)
Within last 6 hours	35.25 (19.96)	35.77 (19.87)

Note: This table reports ED patient arrival volume for physicians with above-median and below-median productivity. Physician productivity is estimated by fixed effects in a regression of length of stay, controlling for all possible interactions of team members (physician assistant or resident and nurse), coworker, and pod; patient characteristics (age, sex, emergency severity index, Elixhauser comorbidity indices); ED arrival volume; and time categories (month-year, day of the week, and hour of the day dummies). The average difference in productivity between physicians of above- and below-median productivity is 0.28, meaning that physicians with above-average productivity have 28% shorter lengths of stay than those with below-average productivity.

Table 1.4.1: Overall Effect of Self-managed System on Log Length of Stay

	(1)	(2)	(3)	(4)	(5)
	Log length of stay				
Self-managed system	-0.095*** (0.012)	-0.108** (0.042)	-0.113*** (0.042)	-0.120*** (0.042)	-0.133*** (0.041)
ED arrival volume and time controls	Y	Y	Y	Y	Y
Pod dummies	Y	Y	Y	Y	Y
Patient demographics	Y	N	Y	Y	Y
Patient clinical information	Y	N	N	Y	Y
Patient triage time	Y	N	N	N	Y
Resident-nurse dummies	Y	N/A	N/A	N/A	N/A
Physician-resident-nurse dummies	N	Y	Y	Y	Y
Number of observations	309,840	310,535	310,535	310,535	310,535
Adjusted <i>R</i> -squared	0.340	0.362	0.368	0.374	0.390
Sample mean log length of stay (log hours)	1.099	1.099	1.099	1.099	1.099

Note: This table reports the effect of the self-managed system on log length of stay, in Equation (1.4.1), while controlling for various observables. All columns control for ED patient arrival volume, time categories (month-year, day of the week, and hour of the day dummies), pod, and patient demographics (age, sex, race, and language). Various models may control for patient clinical information (Elixhauser comorbidity indices, emergency severity index); the time spent in triage, which reflects the triage nurse's subjective belief about patient severity; and resident-nurse interactions or physician-resident-nurse interactions. All models are also clustered by physician. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 1.4.2: Overall Effect of Self-managed System on Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	30-day mortality	Hospital admissions	14-day bounce- backs	Relative Value Units (RVUs)	Log total costs	Log time to first order
Self-managed system	0.0019 (0.0029)	0.0004 (0.0086)	-0.0122* (0.0067)	-0.015 (0.028)	-0.016 (0.030)	-0.026 (0.024)
Number of observations	312,800	312,800	312,800	251,273	280,997	297,693
Adjusted <i>R</i> -squared	0.298	0.464	0.024	0.404	0.529	0.161
Sample mean outcome	0.020	0.272	0.067	2.701	6.724	-0.624

Note: This table reports the effect of the self-managed system on outcomes other than length of stay, estimated by Equation (1.4.1). Fourteen-day bounce-backs are defined as patients who return to the ED within 14 days of being discharged home. Relative Value Units (RVUs) represent intensity of care and are directly scalable to dollar amounts of clinical revenue. Total costs include direct costs for the entire visit, which may include hospital admission. Time to first order is the time between patient arrival at the pod and the first physician order, measured in log hours. All models control for ED patient arrival volume, time categories (hour of the day, day of the week, and month-year dummies), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. All models are clustered by physician. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 1.5.1: Foot-dragging as Expected Future Work Increases

	(1)	(2)	(3)	(4)
	Log length of stay			
Measure of expected future work	ED arrival volume	ED arrival volume	Waiting census	Waiting census
Expected future work	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)
Expected future work \times self-managed system	-0.006*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)
Self-managed system	0.037 (0.044)	0.032 (0.044)	-0.004 (0.039)	-0.010 (0.039)
Pod-specific volume control	N	Y	N	Y
Number of observations	282,105	282,105	282,105	282,105
Adjusted <i>R</i> -squared	0.439	0.440	0.439	0.440
Sample mean log length of stay (log hours)	1.179	1.179	1.181	1.181
Sample mean patient volume measure	15.58	15.58	8.78	8.78

Note: This table shows the effect of expected future work on log length of stay, estimated by Equation (1.5.1). Expected future work is measured either as the number of patients arriving at ED triage during the hour prior to the index patient's arrival at the pod ("ED arrival volume") or as the number of patients in the waiting room during that time ("waiting census"). Models (1) and (3) do not control for pod-level prior patient volume, defined as the number of patients arriving in the pod of the index patient one, three, and six hours prior to the index patient's arrival, while models (2) and (4) do. All models control for time categories (month-year, day of the week, and hour of the day dummies), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), and physician-resident-nurse interactions. All models are clustered by physician. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 1.5.2: Effect of Expected Future Work on Other Outcome and Process Measures

	Outcomes				Process Measures			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	30-day mortality	Hospital Admissions	14-day bounce- backs	Log total costs	Total order count	Laboratory order count	Medication order count	CT or MRI ordered
ED volume	-0.0001 (0.0001)	-0.0006* (0.0004)	-0.0002 (0.0002)	-0.0022* (0.0012)	-0.0117* (0.0071)	-0.0041 (0.0042)	-0.0052** (0.0023)	0.0002 (0.0003)
ED volume × self-managed	0.0002 (0.0003)	-0.0006 (0.0004)	0.0001 (0.0003)	-0.0011 (0.0027)	-0.0094 (0.0101)	-0.0027 (0.0060)	-0.0024 (0.0033)	-0.0002 (0.0005)
Self-managed	-0.001 (0.006)	0.004 (0.012)	-0.011 (0.009)	-0.0062 (0.0642)	-0.526* (0.280)	-0.060 (0.167)	-0.109 (0.090)	-0.021 (0.013)
Adjusted <i>R</i> -squared	0.338	0.461	-0.038	0.526	0.548	0.489	0.402	0.295
Sample mean outcome	0.019	0.265	0.066	6.684	13.267	5.337	2.688	0.246

Note: This table shows the effect of expected future work on other outcomes and process measures, estimated by Equation (1.5.1). Log total costs include all direct costs incurred from the encounter, including any from admissions. Expected future work is measured by ED arrival volume (“ED volume” for brevity), defined as the number of patients arriving at ED triage during the hour prior to the index patient’s arrival at the pod. Models do not include pod-level prior patient volume; results are unchanged when including this. All models control for time categories (month-year, day of the week, and hour of the day dummies), pod, patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. Results are insensitive to controlling for pod-level patient volume. All models are clustered by physician. All models have 289,132 observations, except for model (4), which has 269,905 observations. The sample mean patient volume is 15.53 for all models, except for the model (4), for which it is 15.48. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 1.6.1: Effect of Peer Presence on Foot-dragging

Sample	(1)	(2)	(3)
	Log length of stay		
	Nurse-managed	Self-managed	Pooled
ED volume	0.006*** (0.001)	0.000 (0.001)	
ED volume \times no peer present	0.024*** (0.003)	0.017*** (0.003)	0.018*** (0.002)
ED volume \times peer present, nurse-managed			0.009*** (0.001)
ED volume \times peer present, self-managed			-0.002** (0.001)
ED volume \times only physician in ED			-0.002 (0.002)
Number of observations	130,581	157,419	296,177
Adjusted <i>R</i> -squared	0.441	0.352	0.365
Sample mean log length of stay (log hours)	0.926	1.179	1.061
Sample mean ED volume	17.118	14.217	15.517

Note: This table reports the effect of expected future work, interacted with peer presence, on log lengths of stay. Expected future work is measured by ED arrival volume (“ED volume” for brevity), defined as the number of patients arriving at ED triage during the hour prior to the index patient’s arrival at the pod. Equation (1.6.1) estimates models (1) and (2). Model (1) is estimated with observations of patients seen by nurse-managed teams; model (2) is estimated with self-managed teams. Both of these models use observations with at least one other physician in the ED, so that foot-dragging entails a negative externality against a current coworker, who may or may not be a peer. Model (3) is estimated by Equation (1.6.2) and includes the full sample. Main effects are included but omitted from the table for brevity. In all columns, the phrase “no peer present” means no other physician in same pod but another physician in ED, while “only physician in ED” means no other physician in entire ED. All models control for time categories (month-year, day of the week, and hour of the day dummies), pod (when applicable), patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. All models are clustered by physician. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 1.6.2: Effect of Peer Relationships on Foot-dragging

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log length of stay			Log length of stay			Log length of stay		
Type of peer	Greater tenure than physician			Above-median productivity			≥ 60 hours prior work together		
ED volume	0.008*** (0.001)	0.003*** (0.001)	0.009*** (0.001)	0.008*** (0.001)	0.004*** (0.001)	0.009*** (0.001)	0.006*** (0.001)	0.004*** (0.001)	0.007*** (0.001)
ED volume × peer type	-0.004*** (0.001)	0.000 (0.002)	-0.003*** (0.001)	-0.002* (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.002 (0.001)	-0.002 (0.001)	0.002* (0.001)
Self-managed × ED volume			-0.008*** (0.001)			-0.007*** (0.001)			-0.005*** (0.001)
Self-managed × ED volume × peer type			0.002 (0.002)			0.001 (0.002)			-0.004* (0.002)
Self-managed × peer type			-0.031** (0.014)			-0.027** (0.013)			0.005 (0.014)
Sample	Nurse- managed	Self- managed	Pooled	Nurse- managed	Self- managed	Pooled	Nurse- managed	Self- managed	Pooled
Adjusted <i>R</i> -squared	0.445	0.364	0.376	0.445	0.364	0.376	0.445	0.364	0.376
Sample mean log length of stay (log hours)	0.936	1.251	1.097	0.926	1.179	1.064	0.926	1.179	1.064
Sample mean ED volume	17.30	15.08	16.16	17.12	14.22	15.53	17.12	14.22	15.53

Note: This table reports effect of expected future work, interacted with a peer type, as by Equation (1.6.3). Expected future work is measured by ED arrival volume (“ED volume” for brevity), defined as the number of patients arriving at ED triage during the hour prior to the index patient’s arrival at the pod. All observations require the presence of a peer. Peers with greater tenure, those with high productivity (faster-than-median fixed effect for lengths of stay), and familiar peers (peers with at least 60 hours of history, approximately the 75th percentile, working with the index physician) are considered. ED volume is demeaned. Coefficients for the direct effect of the peer type and pooled coefficients for self-managed are omitted for brevity. All models control for time categories (month-year, day of the week, and hour of the day dummies), pod (when applicable), patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. The nurse-managed, self-managed, and pooled samples had 121,024, 126,264, and 247,288 observations, respectively. All models are clustered by physician. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 1.6.3: Foot-dragging Depending on Physician and Peer Censuses

Peer census quintile	Nurse-managed					Self-managed				
	Index physician census quintile					Index physician census quintile				
	1	2	3	4	5	1	2	3	4	5
1	0.040*** (0.005)	0.030*** (0.004)	0.024*** (0.005)	0.019*** (0.005)	0.009 (0.006)	0.008 (0.008)	0.003 (0.007)	0.008* (0.004)	0.008 (0.005)	0.006 (0.006)
2	0.023*** (0.006)	0.020*** (0.006)	0.018*** (0.004)	0.017*** (0.003)	0.005 (0.004)	0.007 (0.012)	0.009 (0.010)	0.012 (0.011)	0.005 (0.007)	0.007 (0.007)
3	0.023*** (0.006)	0.009 (0.007)	0.001 (0.006)	0.009** (0.004)	0.002 (0.005)	0.006 (0.008)	0.010 (0.007)	0.010 (0.011)	0.008 (0.010)	0.006 (0.016)
4	0.022*** (0.004)	0.010** (0.004)	0.001 (0.004)	0.002 (0.004)	-0.006 (0.004)	0.003 (0.007)	0.008 (0.005)	0.012 (0.007)	0.011 (0.008)	-0.004 (0.014)
5	0.010** (0.004)	0.002 (0.005)	-0.007* (0.004)	0.002 (0.005)	0.000 (0.009)	-0.002 (0.007)	0.000 (0.009)	0.006 (0.012)	0.002 (0.014)	0.004 (0.040)

Note: This table reports the effect of expected future work, for each cell of physician and peer quintile interaction and for each organizational structure. Expected future work is measured by ED arrival volume, defined as number of patients arriving at ED triage during the hour prior to the index patient's arrival at the pod. All observations involve at least one peer present. On the left are coefficients $\alpha_1^{m,n}$ from Equation (1.6.4) for physician-census quintile m and peer-census quintile n using nurse-managed-team observations (adjusted R -squared 0.428, 121,024 observations). On the right are the same coefficients using self-managed-team observations (adjusted R -squared 0.329, 126,264 observations). All models control for time categories (month-year, day of the week, and hour of the day dummies), pod (when applicable), patient demographics (age, sex, race, and language), patient clinical information (Elixhauser comorbidity indices, emergency severity index), triage time, and physician-resident-nurse interactions. All models are clustered by physician. Coefficients for the nurse-managed team are also shown in Appendix Figure 1.A.5. * significant at 10%; ** significant at 5%; *** significant at 1%.

1.A Empirical Appendix

1.A.1 Exogenous Variation in Patients and Physicians

1.A.1.1 Similar Exposures across Physician Types

I first show similar characteristics for patients available to physicians with different preferences and productivity (or speed). Preferences for specific patient types are estimated by the probability that a physician will choose a patient type when given the choice. Productivity measures are estimated using fixed effects for physician identities in a regression of log length of stay. Physicians with one standard greater preferences for a patient type are 7.4% more likely to choose that patient type than average. Physicians who are one standard deviation faster than average have 11% shorter lengths of stay.

Regardless of these physician types, I find that the patients that they are exposed to are roughly equal regardless of them. In Table 1.2.1 in the main chapter, I show that physicians who are faster or slower than median are exposed to the same patient types. In Table 1.A.1, I show that in the self-managed team, patients entering the pod, whether it be a pod with a self-managed team or a nurse-managed team, are the same regardless of preferences of the physicians working in the pod.

Also in Table 1.2.1 in the main chapter and 1.A.1 in this appendix, I also show that physicians of different types are exposed to similar patient volumes arriving to the ED. Specifically, higher-productivity physicians are equally likely as lower-productivity ones to work during times with high patient volume. Table 1.2.2 in the main chapter shows this result by average patient volumes for each higher- or lower-productivity physician type, and Figure 1.A.1 presents this nonparametrically by plotting the distribution of patient volume for each of these types.

1.A.1.2 Joint Insignificance of Physician Identities

In order to show formally conditional randomization of physicians to shifts, I test for the joint significance of physician identities in regressions of patient characteristics arriving at the pod and ED patient volume, while conditioning on rough indicators of time.

For patient characteristics, I first summarize patient characteristics of age, sex, emergency severity index (ESI), race, and language into a propensity score of length of stay. I calculate this propensity score for each patient and then average these propensity scores for patients arriving at each pod and hour-date in my data. Using physician schedules, I associate the average propensity score for each pod-hour combination to physicians that are working on that pod and during that hour-date. I then estimate this equation:

$$AvgProp_{pt} = \alpha_j I_{jpt} + \eta_{pt} + \varepsilon_{jpt},$$

where $AvgProp_{pt}$ is the average patient propensity score for pod p at time (hour-date combination) t , I_{jpt} is an indicator variable equal to 1 if physician j is working in pod p at time t , and η_{pt} is a

sum of fixed effects for interactions between pod p and rough time dummies including month-year dummies, day of the week dummies, and hour of the day dummies. I test for the joint significance of the vector of coefficients $\alpha = (\alpha_j)$. The F -statistic (64, 659042) under the null that $\alpha = \mathbf{0}$ is 1.07 (p -value of 0.32), clustering by date.

For ED volume, I perform a similar exercise. I estimate this equation:

$$EDWork_t = \alpha_j I_{jt} + \eta_t + \varepsilon_{jt},$$

where $EDWork_t$ is the number of patients arriving at the ED during time (hour-date combination) t , I_{jt} is an indicator variable equal to 1 if physician j is working in the ED at time t , and η_t is a sum of fixed effects for rough time categories including month-year, day of the week, and hour of the day. The F -statistic (64, 2098) under the null that $\alpha = \mathbf{0}$ is 1.11 (p -value of 0.26) when clustering by unique date (each of the 75 physicians in the sample works from 60 to 845 days).

1.A.1.3 Exogenous Assignment of Physicians to Peers

I also show that physician identities do not explain the preferences or ability of their peers that they happen to be working with on a shift. With respect to physician productivity, I regress the productivity (length of stay) fixed effect of the physician against that of his peer. That is, if $\hat{\alpha}_j$ is the productivity fixed effect for physician j , and $-j(t)$ denotes physician j 's peer during shift t , I perform this regression:

$$\hat{\alpha}_{-j(t)} = \beta \hat{\alpha}_j + \varepsilon_{jt}.$$

I estimate the coefficient β to be small and insignificant at -0.003 with a standard error of 0.014 (p -value of 0.84). Similarly, the correlation coefficient between physician and peer fixed effects is -0.018 (p -value 0.16).

If I normalize the average productivity fixed effect of peers working with physicians who are faster than average to be 0 (with standard deviation 0.107), the average productivity fixed effect of peers working with physicians who are slower than average is 0.0001 with standard deviation of 0.105.

If instead of entering the physician's productivity fixed effect on the right-hand-side, I enter physician dummies in the specification

$$\hat{\alpha}_{-j(t)} = \eta_j + \varepsilon_{jt},$$

where the error term is clustered by unique pairings of physician and peer, I am unable to reject the null that the vector η is jointly 0 with an F -statistic (53, 1995) of 0.57 (p -value 0.99).

I perform similar analyses with respect to physician preferences and find no relationship between the preferences of peers.

1.A.2 Inference of Overall Effect

In Section 1.4 in the main chapter, I estimate the overall effect of self-managed teams as a benchmark that includes all potential mechanisms, including foot-dragging. I include pod fixed effects, since Bravo changed organizational systems from nurse-managed to self-managed. I also control for provider identities, as I observe the same providers in both pods over time. I finally control for rich patient characteristics.

In Section 1.A.2.1, I show unadjusted outcomes in order to clarify patient assignment to the two pods in each month. In Sections 1.A.2.2 and 1.A.2.3, I discuss two alternative methods of inference that allow for pod-level shocks, given that I have only two pods and cannot cluster by pod for inference.

1.A.2.1 Patient Assignment and Unadjusted Outcomes

I examine the assignment of patients to each pod. Table 1.A.2 presents average patient characteristics for patients in Alpha and in Bravo. It shows that older patients and patients with more severe conditions (a lower emergency severity index indicates a more severe condition) were generally sent to Alpha.

I next examine the expected length of stay of patients, based only on their characteristics, in each pod over time. I first estimate the following regression, using only visits to Alpha pod in 2005:

$$Y_{ijkt} = \beta X_{it} + \nu_{jk} + \varepsilon_{ijkt}, \quad (1.A.1)$$

where as before X_{it} are patient characteristics (age, sex, race, language, emergency severity index) for patient i at visit arrival time t , and ν_{jk} are absorbing fixed effects for the provider trio. I then use estimates $\hat{\beta}$ from Equation (1.A.1) to generate expected lengths of stay for all patients, ignoring the provider trio this time. Figure 1.A.2 shows the plot of average patient expected length of stay based on patient characteristics, for both pods and each month. Patients with longer expected lengths of stay are always sent to Alpha, but this differential reduces over time.

I finally plot out average, unadjusted lengths of stay for both pods and each month in Figure 1.A.3. These unadjusted numbers show a gradually increasing trend in length of stay in Bravo and also a decreasing trend in Alpha. These trends occur prior to the Bravo regime change in March 2010. Also, Alpha's trend is continuous across the regime change, while lengths of stay in Bravo appear to have a sudden drop at the regime change, followed by a continued trend afterward. As suggested by Figure 1.A.2, part of this is due to more intensive patients (i.e., patients expected to stay longer) being assigned to Bravo over time. Another reason for this is that new nurses spent more time in Bravo as it expanded, and these new nurses were less efficient than older, more experienced nurses. While these trends are not unconditionally parallel, I show in Section 1.4 and Figure 1.4.1 in the main chapter that they are quite parallel when controlling for patient characteristics and provider identities.

1.A.2.2 Inference with Serially Correlated Pod-level Error Terms

In this approach, I address sampling variation at the pod level across time but with a more parametric form on the error terms. Specifically, I allow for pod-month random shocks that are serially correlated across months by a first-order autoregressive (AR1) process. That is, I first calculate month-year-pod fixed effects from

$$Y_{ijkpt} = \sum_{m=1}^M \sum_{y=1}^Y \alpha_{myp} I_{t \in m} I_{t \in y} + \beta X_{it} + \tilde{\eta}_t + \nu_{jk} + \varepsilon_{ijkpt}, \quad (1.A.2)$$

which is the same model used in the main chapter to generate Figure 1.4.1. $I_{t \in m}$ and $I_{t \in y}$ are indicator functions for t belonging in month m and year y , respectively, and $\tilde{\eta}_t$ is a revised sum of fixed effects only for day of the week and hour of the day. Coefficients estimated for α_{myp} are used as data points.

In the second stage, I estimate a model with observations collapsed to the month-year-pod level:

$$\hat{\alpha}_{myp} = \gamma \text{Self}_{myp} + \eta_{my} + \zeta_p + \varepsilon_{myp},$$

where $\hat{\alpha}_{myp}$ are estimated coefficients from Equation (1.A.2), γ is the coefficient of interest on the treatment indicator Self_{myp} of whether pod p is a self-managed system during month m and year y , η_{my} and ζ_p are respective time and pod fixed effects, and ε_{myp} is a serially correlated error term with the AR1 process

$$\varepsilon_{myp} = \rho \varepsilon_{my-1,p} + z_{myp}.$$

I estimate a self-managed effect of $\hat{\gamma} = -0.0981$, with a standard error of 0.0281, which is significant with a p -value of 0.001. This estimate is quite similar to the baseline estimate in the main chapter. The estimated correlation in error terms across months is $\hat{\rho} = 0.302$.

1.A.2.3 Inference with Systematic Placebo Tests (Randomization Inference)

Another alternative method of inference takes sampling as fixed and instead considers randomization at the level of treatment (Rosenbaum, 2002). This is also in the same spirit of systematic placebo tests (e.g., Abadie, Diamond, and Hainmueller, 2011; Abadie, 2010). Under the sharp null of no effect of the self-managed system, there should be no significant difference between my obtained estimates and those I would obtain if I consider a number of placebo regime changes over each pod and month.

Here, I again use estimated month-year-pod fixed effects $\hat{\alpha}_{myp}$ from Equation (1.A.2). I then perform a regression discontinuity at placebo regime changes at each month and pod with a bandwidth of three months prior to the placebo regime change date and three months post. That is,

using only only observations from within the bandwidth for the two pods, I estimate

$$\hat{\alpha}_{myp} = \gamma^r PlaceboSelf_{myp}^r + \eta_{my} + \zeta_p + \varepsilon_{myp},$$

where $PlaceboSelf_{myp}^r$ is an indicator for whether pod p at month m and year y is self-managed under the placebo regime change r . Out of the 130 estimates for γ^r , the true regime change had the largest coefficient, corresponding to a randomization-inference p -value of about 0.008.

1.A.3 Other Behavioral Mechanisms

In this section, I examine other behavioral mechanisms that suggest social incentives and strategic behavior. These are not central to the overall difference between the nurse-managed and self-managed systems. In the first subsection, I examine general peer effects akin to those studied in Mas and Moretti (2009). Rather than considering peer effects on foot-dragging, which is an interaction of the peer and expected future work, as in the main chapter, I consider the direct effect of peer. Consistent with Mas and Moretti (2009), I find that working with a faster peer shortens lengths of stay for the index physician, but I do not find a significant difference in this peer effect between organizational systems. This suggests that social incentives exist in both systems.

The second subsection considers physician behavior at the end of shift. I find that physicians discharge patients substantially sooner when nearing the end of their shifts. This could be consistent with moral hazard, either foot-dragging when not at the end of shift or prematurely discharging patients at the end of shift, but does not substantially differ between organizational systems. I also find some evidence increased free-riding at the end of shift, although this is minor relative to the overall effect, as I show in the main chapter.

1.A.3.1 Peer Effects

Mas and Moretti (2009) motivate their paper by the fact that, in team production, individual output is difficult to observe by managers. If workers only care about the effort of work and the chance of being fired if they are suspected of shirking, then they should rationally reduce their work if they are working with someone who makes the team more productive. That is, working with a productive teammate should have negative “peer effects” on the performance of physicians. On the other hand, if they care about what their peers think about them or if they are altruistic towards peers, then there will be social incentives for increasing productivity when working with a productive peer (Fehr and Gächter, 2002; Kandel and Lazear, 1992).⁴⁵

In our setting, however, it is easy to see that strategic behavior in the absence of any social incentives may also lead to positive peer effects. A physician working with a less-productive peer will be more likely to get new work unless if he slows down. Similarly, if there is a cost to

⁴⁵Mas and Moretti (2009) discuss other mechanisms by which there may be positive peer effects, including that less-productive workers may learn from more-productive ones when together.

keeping patients longer than necessary (e.g., patients become unhappy, or physicians become bored), then working with a more-productive peer could possibly speed up the physician. Moreover, other mechanisms discussed in the main chapter could influence the sign and magnitude of peer effects. Free riding by waiting for productive peers to choose work would have a negative influence on peer effects, while dynamic smoothing and matching may have a positive influence if productive peers have complementary skills and availability or a negative influence if less-productive peers have the complementary skills and availability.

Finally, social incentives may differ between self-managed teams and nurse-managed teams. In nurse-managed teams, peers impose a negative externality by being less productive (i.e., the foot dragging externality), and social incentives may lead increased efficiency. However, in self-managed teams, peers on the same team actually may impose a positive externality by being less productive to others within the team, because they prevent work from being sent to the pod, although they impose a negative externality to other teams. In addition to the direction of the social incentives, their strength may differ between the two organizational systems as peers work more closely in self-managed teams.

Although peer effects may be difficult to interpret in terms of mechanisms for these reasons, they may still be useful to compare between the two organizational systems in reduced form. It is certainly possible that some physician-peer combinations may perform better in a self-managed setting while others may perform better in a nurse-managed setting. Employing similar methodology as Mas and Moretti (2009), I first estimate physician fixed effects for log length of stay, and then I use the fixed effects of peers as an explanatory variable in a regression of productivity in order to estimate peer effects.

In the first stage, I estimate the following regression on log length of stay Y_{ijkpt} for patient i , physician j , resident-nurse team k , and visit arrival time t :

$$Y_{ijkpt} = \theta_j + \mathbf{M}\phi_{C_j} + \beta X_{it} + \zeta_p + \eta_t + \varepsilon_{ijkpt}, \quad (1.A.3)$$

where in addition to patient characteristics X_{it} and fixed effects for time categories η_t , I control for all possible sets of physicians j , resident-nurse combinations k , physician peers l , and pod locations p in order to address the reflection problem (Manski, 1993). These providers and peers are included in a set of all dummies representing each possible combinations $\phi_{C_j} = \{C(j, k, l, p)\}$, used in estimation with a vector \mathbf{M} of nuisance parameters, where

$$C(j, k, l, p) = \begin{cases} 1 & \text{if physician } j \text{ is working with team } k \text{ and peer } l \text{ in pod } p, \\ 0 & \text{if } i = l, \\ 0 & \text{otherwise.} \end{cases}$$

The parameter of interest is the physician fixed effect θ_j . The standard deviation of estimated fixed effects is 0.11, meaning that physicians one standard deviation above mean productivity have

lengths of stay that are 11% shorter than average.

In the second stage, I use the set of fixed effects estimated in (1.A.3) in this regression, using the fixed effect θ_{-j} for physician j 's peer:

$$Y_{ijkpt} = \alpha\theta_{-j} + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}. \quad (1.A.4)$$

The coefficient α represents peer effects from working with a peer with productivity θ_{-j} . A positive γ suggests that physicians work faster when working with a more-productive (faster) peer and slower when working with a less-productive (slower) peer. I estimate (1.A.4) using observations from nurse-managed teams and self-managed teams separately. I also estimate a pooled version with

$$Y_{ijkpt} = \alpha_1\theta_{-j} \cdot Self_{it} + \alpha_2\theta_{-j} + \alpha_3Self_{it} + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}.$$

All regressions require that there be a peer present in the pod.

Table 1.A.3 reports estimates of peer effects. A 1% increase in peer productivity leads to a 0.1% increase in physician productivity. Faster peers have a stronger influence than slower peers in both settings. However, peer effect estimates can be relatively imprecise, compared to foot-dragging results in the main chapter and in particular in the self-managed system. I cannot reject that overall peer effects are different between the two organizational systems.

Unlike foot-dragging in the main chapter, the peer effects estimated in Table 1.A.3 could represent a number of different mechanisms. While effects are quantitatively similar in self-managed teams, they are less precisely estimated than in nurse-managed teams, despite similar numbers of observations. This suggests that peer effects and perhaps the interaction between physicians in self-managed teams, through any mechanism, are generally less predictable. Although I do not show results here, I do not find any significant effect of peer effects of working with a peer with productivity specific to the index patient. That is, working with a physician who is better at seeing heart patients does not improve the productivity of the index physician seeing the heart patient. In addition, I do not find peer effects on quality outcomes. These facts suggests that peer effects do not occur by learning or being helped by more skilled physicians, which Mas and Moretti (2009) also rule out.

1.A.3.2 End of Shift Behavior

An important institutional fact is that physicians in the ED are responsible for the entire care of patients that they choose and usually do not hand off their patients when ending their shifts.⁴⁶ Using this fact, I then study the behavior of physicians nearing the end of their shift. The desire for physicians not to stay past the end of their shift can manifest in two types of moral hazard that are

⁴⁶In the vast majority of cases, physicians complete work on their patients (and discharge their patients) before leaving the ED. In a few cases, they may hand off a patient to a colleague if the patient is awaiting results or disposition that is expected to take a long time and if work is completed for the near future.

similar to those corresponding to the desire to avoid future patients. First, physicians may free-ride in the self-managed system, waiting for their peer to choose the patient (or passing the patient onto the next physician). Although in the main chapter I show that it does not account for a significant part of the overall effect between organizational systems, its relative importance could be greater at the end of shift, even if it is not substantial on average.

The second relevant moral hazard is foot-dragging. Recall that foot-dragging delays the discharge of patients in order to avoid future work. However, when physicians near the end of their shift, there should be less future work to avoid. In the extreme, if the physician is about to end his shift in the next minute, then it is very unlikely that new work will be assigned to him for the remainder of his shift, both because of the short time span and because the triage nurse (in the nurse-managed system) is unlikely to send new patients to physicians who are ending their shifts. Note that there may be an opposite moral hazard at play at the end of shift: Physicians may want to discharge their patients earlier than optimal in order to get home sooner.

In order to show that physicians have behavioral scope for the time of discharge, I first plot a kernel density of discharge order times relative to the end of the physician's end of shift. I show this in Figure 1.A.4. Consistent with the second moral hazard discussed above, physicians are more likely to discharge their patients right as they end their shift. In addition to showing evidence of moral hazard, it also shows the more basic fact that physicians can and do manipulate the discharge times of their patients, which is necessary for the foot-dragging moral hazard shown in the main chapter.

Next, I evaluate the effect of patient arrival near the end of a physician's shift with the following regression:

$$Y_{ijkpt} = \alpha_1 Self_{pt} \cdot EndShift_{it} + \alpha_2 EndShift_{it} + \alpha_3 Self_{pt} + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt}. \quad (1.A.5)$$

$EndShift_{it}$ is an indicator variable for whether the patient arrives within two hours of the end of a physician's shift. In the nurse-managed system, the relevant physician is the one assigned the patient; in the self-managed system, I consider the end of shift of either of the two physicians who could choose the patient. As before, $Self_{it}$ is an indicator for self-managed teams, X_{it} are patient characteristics, ζ_p is a pod fixed effect, η_t is a sum of fixed effects for time categories, and ν_{jk} are provider-trio fixed effects. The coefficients of interest are α_1 and α_2 . In addition, as in the foot dragging regressions, α_3 represents the residual effect of self-managed teams after accounting for free riding at the end of shift.

In the first column of Table 1.A.4, I show results for log length of stay. Because this duration is defined as the time from pod arrival at discharge order, it includes the time required for a physician to sign up for the patient in the self-managed system. I find a significant decrease in both lengths of stay in the nurse-managed system for physicians assigned a patient near the end of their shift.

In in the nurse-managed system, the length of stay for such patients is 38% lower, compared to patients arriving before the last two hours of the shift. The interaction between end of shift and the self-managed system has a coefficient of equally large magnitude in the opposite direction. The main effect of self-managed system is relatively unchanged compared to results shown in Table 1.4.1 in the main chapter.

Note that, for the self-managed system, I do not condition on the physician who treats the patient being at the end of his shift in these baseline results because this would be similar conditioning on a dependent variable. However, I find that physicians who see patients at the end of their shift in the self-managed system discharge their patients similarly early relative to patients not at the end of their shifts. This does not appear to be because they choose patients who are likely to be discharged in a short time. If anything, they are more likely to choose sicker patients at the end of their shift, perhaps because these patients require urgent attention. Therefore, the greater length of stay for patients arriving at the end of shift for any physician in the self-managed system relative to those arriving near the end of shift in the nurse-managed system appears to represent free-riding. Physicians nearing the end of their shift are unlikely to choose new patients in the self-managed system, because their peer will likely pick that patient.

In the second column of Table 1.A.4, I show results for log time to first order. As noted in the main chapter, log time to first order is generally very small, implying about half an hour on average, and so effects are generally not large in absolute value. However, I do find a similar relative decrease in the log time to first order in the nurse-managed system for patients arriving near the end of shift, and a similar equally large increase corresponding to the interaction of end of shift and the self-managed system. This again implies that physicians pay attention more quickly to patients when they are assigned patients near the end of their shifts in the nurse-managed system, but patients are not attended to any more quickly when arriving near the end of shift of a physician who *could* choose them in the self-managed system. Again, this is because physicians nearing the end of their shift in the self-managed system are unlikely to choose new patients, reflecting free-riding at the end of the shift.

The question remains whether patients being discharged more quickly when assigned to physicians near the end of their shift mostly represents a reversal of foot-dragging or a new moral hazard in which physicians discharge patients sooner than appropriate because they want to go home. Of course, the effect may be a mixture of the two mechanisms. In Table 1.A.5, I attempt to shed light on this by examining the effect on other outcomes of quality and utilization for patients arriving near the end of shift. I find no effect on quality measures of 30-mortality, 14-day bounce-backs, and hospital admissions. For the latter outcome, the effect is fairly tightly estimated at -0.53 (95% CI -2.1 to 1.0%) compared to the baseline of 27.2%. I do find somewhat statistically significant effects for Relative Value Units (RVUs), which represent revenue collected by the ED due to physician intensity of care, and log total costs. However, these effects are generally small compared with the large 40% effect on length of stay: Physicians incur 7% fewer RVUs and 5% fewer costs for patients

arriving at the end of shift, when comparing the nurse-managed system with the self-managed system. Recall that patients arriving at “the end of shift” in the self-managed system may not be cared for by the physician nearing his end of shift.

The at best small magnitudes of effect on quality and utilization could be interpreted as support for the reversal of foot-dragging, rather than physicians providing insufficient care, as the primary behavioral mechanism for physicians near the end of shifts. This would imply that the foot-dragging identified in the main chapter by increases in length of stay with expected future work are a gross lower bound. That is, physicians keep patients longer than medically necessary, in the self-managed system and even when expected future work is close to none. However, this would not be consistent with theory in which pure foot-dragging is the delay in discharge in order to avoid future work. Instead, grossly longer lengths of stay for patients not at the end of shift could be consistent with physician risk aversion, which may not be in line with those of the ED, triage nurse, or the patient.

1.A.4 Additional Results

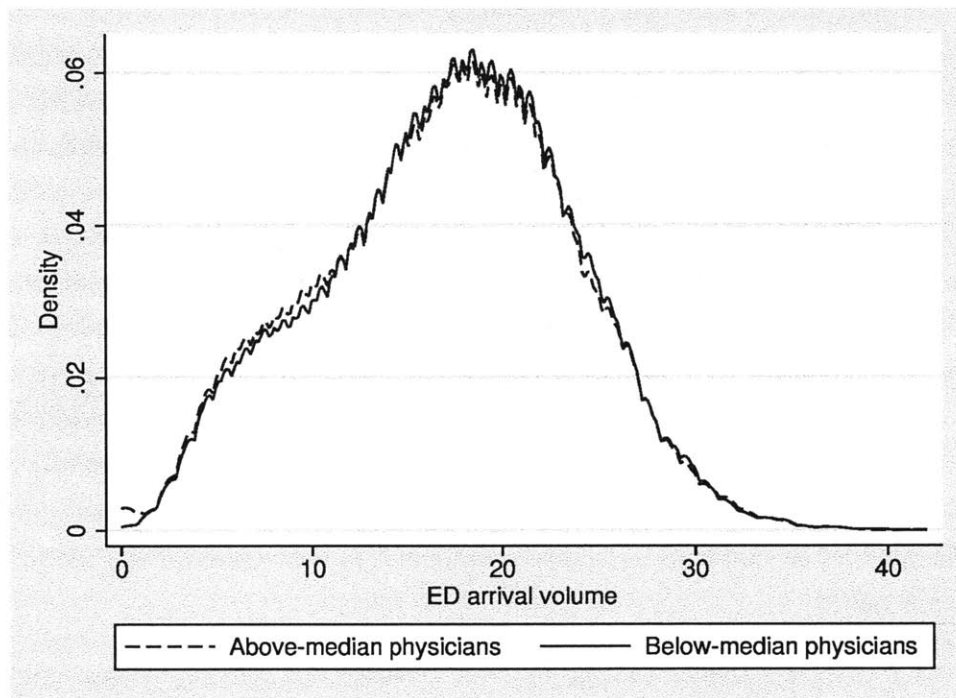
In this Appendix, I present the following additional empirical results, as well as a brief discussion of each set of results:

- Figure 1.A.5: Presents a graphical representation of foot-dragging as a census policy, corresponding to Table 1.6.3 in the main text.
- Figure 1.A.6: Presents the correlation between censuses and new patient assignment in both pods over time, with confidence intervals on estimates in Bravo pod (analogous to Figure 1.7.2 in the main text).
- Figure 1.A.7: Presents the correlation between censuses and new patient assignment in both pods over time, with confidence intervals on estimates in Alpha pod (analogous to Figure 1.7.2 in the main text).

Figure 1.A.5 graphically presents the same results as in Table 1.6.3 in the main text. It shows that foot-dragging decreases with increasing censuses both for the index physician and the peer. As also shown in Table 1.6.3, there is a depression in foot-dragging when both physicians have censuses in the third quintile. As discussed in the main text, this could be consistent with social preferences, in particular an aversion for inequality.

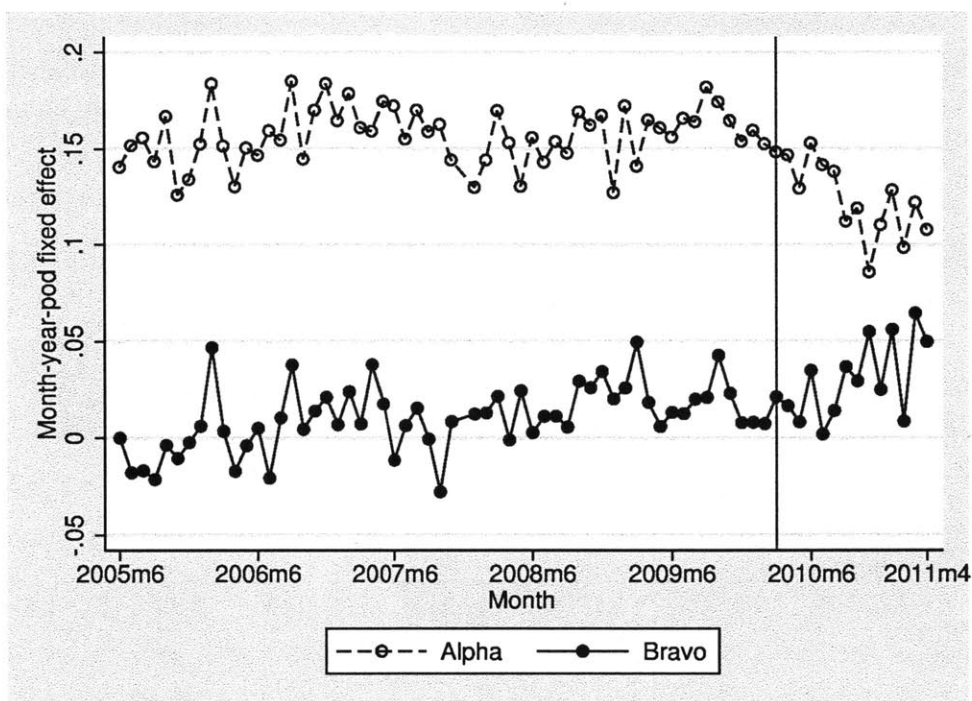
Figures 1.A.6 and 1.A.7 present the same results as in Figure 1.7.2 in the main text, plotting the coefficient on censuses in a linear probability model of new patient assignment. The only difference between these two Appendix Figures and Figure 1.7.2 is that confidence intervals are plotted for Bravo and Alpha pods, respectively. For most months, I can reject that the central estimate for one pod is the same as the central estimate for the other pod, but the two confidence intervals otherwise overlap. Of course, if I widen the kernel interval, then I can narrow the confidence intervals. Also, all central estimates for Alpha pod are below those for Bravo pod in the pre-regime period.

Figure 1.A.1: Density of Patient Volume to ED for High- and Low-Productivity Physicians



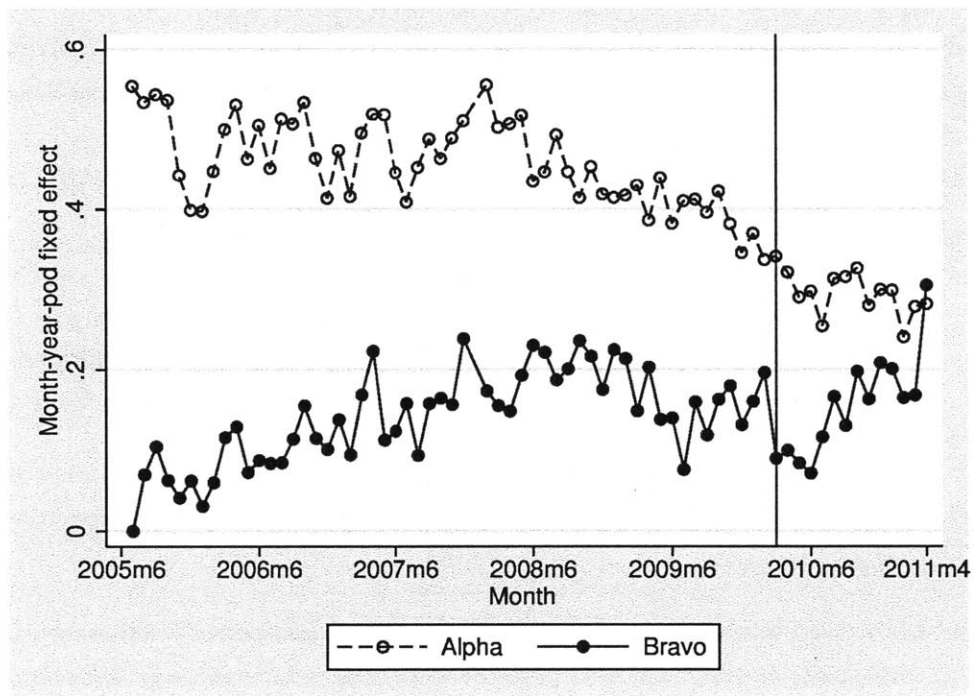
Note: This figure shows a kernel density plot of the ED arrival volume (number of patients arriving at the ED) during each hour while physicians of above- and below-median productivity are working. The density for above-median physicians is shown in dashed blue; the density for below-median physicians is shown in solid red. Physician productivity is estimated by fixed effects in a regression of length of stay, controlling for all possible interactions of team members (physician assistant or resident and nurse), coworker, pod location; patient demographics (age, sex, emergency severity index, Elixhauser comorbidities); ED arrival volume; and time dummies (month-year combination, day of the week, and hour of the day). The average difference in productivity between physicians of above- and below-median productivity is 0.28, meaning that physicians with above-average productivity take 28% less time than those with below-average productivity.

Figure 1.A.2: Relative Predicted Log Length of Stay from Patient Characteristics



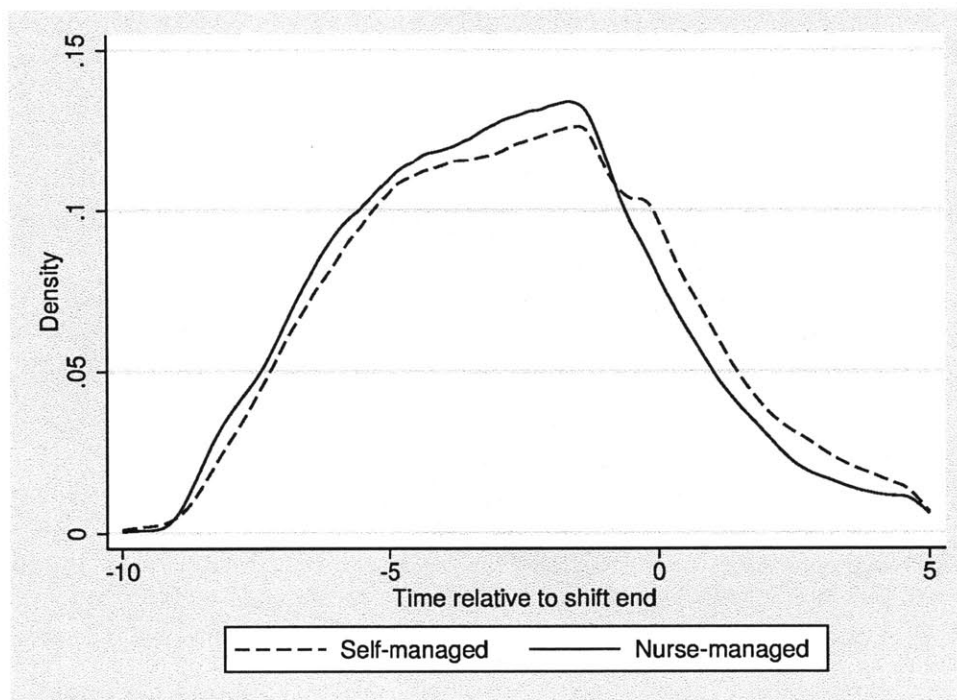
Note: This figure shows average predicted log length of stay for patients in each pod, month, and year, where the average is normalized to 0 for Bravo pod at the beginning of the panel. In the first step, Equation (1.A.1) is estimated for patients going to Alpha pod in 2005. Coefficients are then used to calculate predicted log lengths of stay for all patients. In the second step, averages for predicted log lengths of stay are computed for each pod, month, and year. Alpha pod averages are plotted with hollow blue circles; Bravo pod averages are plotted with solid red circles. The vertical red line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure 1.A.3: Unadjusted Average Log Length of Stay by Pod and Month-year



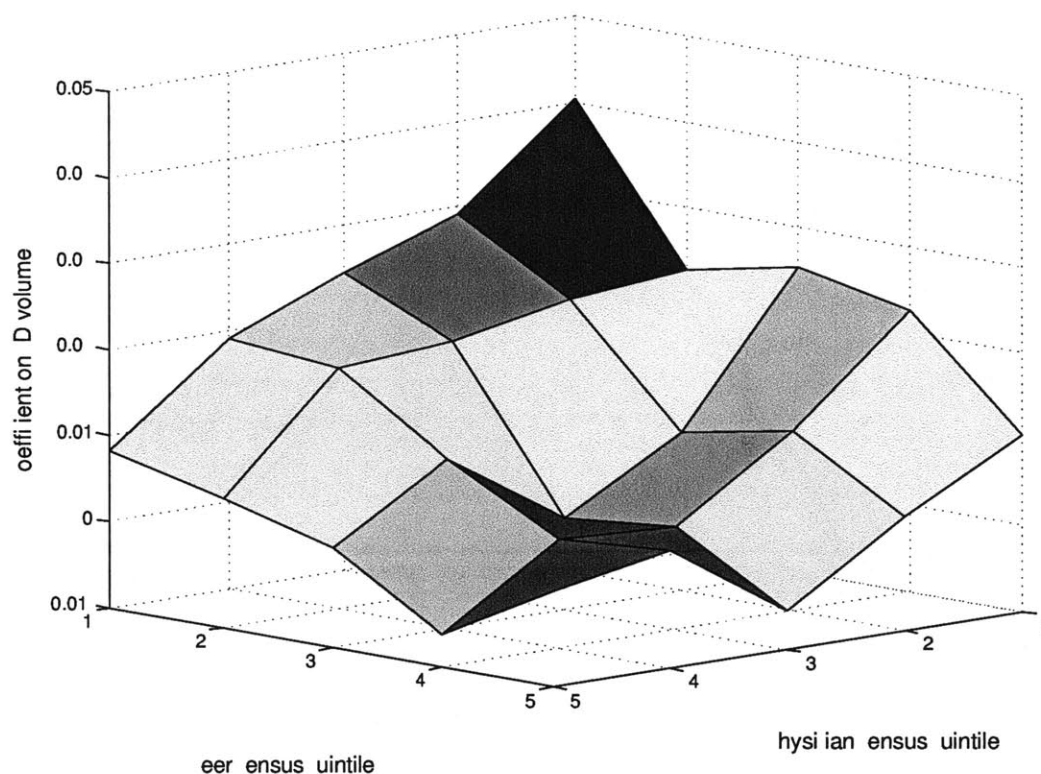
Note: This figure shows average log length of stay for each pod, month, and year, where the average is normalized to 0 for Bravo pod at the beginning of the panel. No patient characteristics or provider identities are controlled for, unlike Figure 1.4.1 in the main chapter. Alpha pod averages are plotted with hollow blue circles; Bravo pod averages are plotted with solid red circles. The vertical red line indicates the month of the regime change of Bravo from a nurse-managed system to a self-managed system, in March 2010. Alpha was always self-managed.

Figure 1.A.4: Kernel Density Plot of Discharge Order Times Relative to Shift End



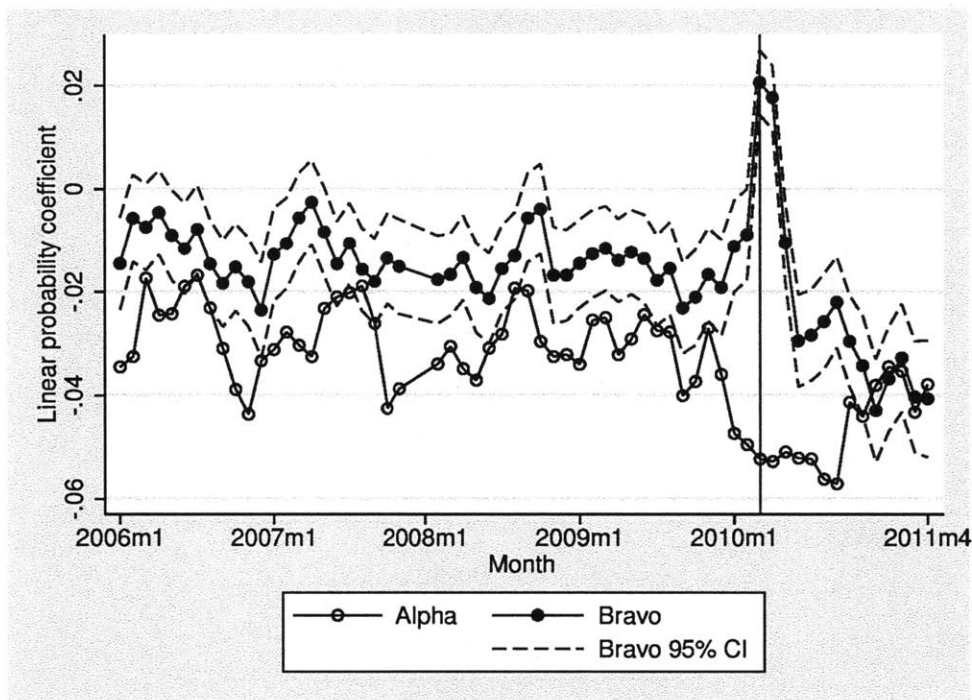
Note: This figure shows the kernel density plot of discharge order times relative to the physician's end of shift. The density for discharge orders in the nurse-managed system is plotted with the solid red line; the density for discharge orders in the self-managed system is plotted with the blue dashed line. Physicians are usually expected to complete care of their patients before going home. Shifts are usually 9 hours in length.

Figure 1.A.5: Foot-dragging as Policy Function of Current Physician and Peer Censuses



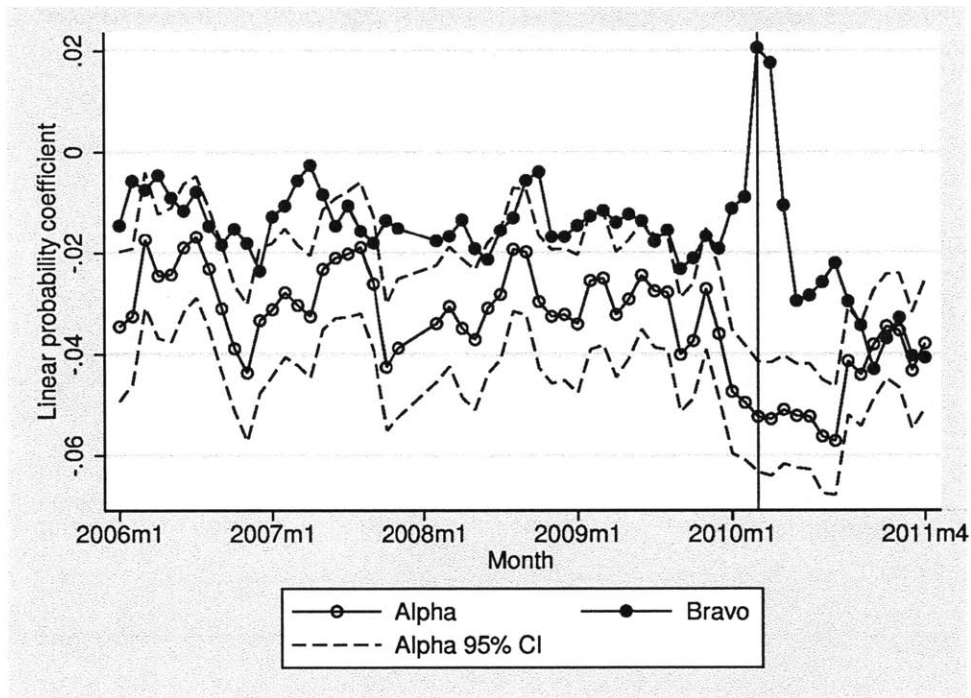
Note: This figure shows foot-dragging, as estimated by the log length of stay coefficient on expected future work (measured by ED patient volume concurrent to the index patient's pod arrival) and as a function of current physician and peer census quintiles. The sample includes observations from the nurse-managed system, in which there is at least one physician peer present in the pod. Actual coefficients, standard errors, and details of the regression that produces this is discussed in Table 1.6.3 in the main chapter.

Figure 1.A.6: Effect of Additional Census Patient on New-patient Assignment Probability



Note: This figure shows the new-patient assignment probability, as a function of relative censuses for physicians within each pod. The plotted coefficient estimates represent the average effect on assignment probability of each additional patient on a physician's census relative to his peer's census. Hollow blue circles show coefficient estimates for Alpha pod, which was always self-managed. Solid red dots show the coefficient estimates for Bravo pod, which switched to a self-managed system in March 2010, shown with a vertical red line. Coefficients are estimated in a kernel regression using a triangular kernel with 45 days on each side; estimates for February and March 2010 in Bravo pod are estimated by a kernel with 45 days only on the same side of the regime change. Confidence intervals are plotted as dashed lines on estimates for Bravo.

Figure 1.A.7: Effect of Additional Census Patient on New-patient Assignment Probability



Note: This figure shows the new-patient assignment probability, as a function of relative censuses for physicians within each pod. The plotted coefficient estimates represent the average effect on assignment probability of each additional patient on a physician's census relative to his peer's census. Hollow blue circles show coefficient estimates for Alpha pod, which was always self-managed. Solid red dots show the coefficient estimates for Bravo pod, which switched to a self-managed system in March 2010, shown with a vertical red line. Coefficients are estimated in a kernel regression using a triangular kernel with 45 days on each side; estimates for February and March 2010 in Bravo pod are estimated by a kernel with 45 days only on the same side of the regime change. Confidence intervals are plotted as dashed lines on estimates for Alpha.

Table 1.A.1: Average Patient Characteristics Available to Physicians with Preference for or against Patient Type

Patient characteristic	Nurse-managed system		Self-managed system	
	Physicians with preference for	Physicians with preference against	Physicians with preference for	Physicians with preference against
Age	43.6 (18.4)	43.8 (18.5)	48.8 (19.6)	48.8 (19.6)
Emergency severity index	3.11 (0.774)	3.12 (0.771)	2.73 (0.773)	2.73 (0.776)
White	0.448 (0.497)	0.445 (0.497)	0.512 (0.500)	0.509 (0.500)
Black or African-American	0.250 (0.433)	0.249 (0.432)	0.233 (0.423)	0.232 (0.422)
Spanish speaking	0.114 (0.318)	0.114 (0.317)	0.097 (0.297)	0.096 (0.294)
Female and age < 35 years	0.266 (0.442)	0.261 (0.440)	0.185 (0.388)	0.184 (0.387)

Note: This table reports average patient characteristics available to physicians with a preference for or against a given patient type, while working in a nurse-managed or self-managed system. In the nurse-managed system, patients are assigned; on the self-managed system, patients are made available by the triage nurse. I construct measures of physician preference by (1) estimating a linear probability model of patient choice using observations in the behavioral system and including physician-specific coefficients on patient characteristics, and (2) selecting physicians with high or low coefficients as having a preference for or against each patient characteristic, respectively. Once I arrive at measures of physician preference for each patient characteristic, I describe the average for that patient characteristic for patients available to choose from for physicians with preferences for or against that characteristic. The emergency severity index (ESI) ranges from 1 to 5, and a lower ESI represents a more severe patient. Standard deviations are reported in parentheses.

Table 1.A.2: Average Patient Characteristics Assigned to Alpha or Bravo Pod

Patient characteristic	Alpha	Bravo
Age	49.9 (19.3)	44.5 (18.8)
Emergency severity index	2.58 (0.725)	3.06 (0.785)
White	0.527 (0.499)	0.458 (0.498)
Black or African-American	0.231 (0.421)	0.247 (0.431)
Spanish speaking	0.086 (0.280)	0.112 (0.315)
Female and age < 35 years	0.269 (0.374)	0.249 (0.432)

Note: This table reports average patient characteristics for patients being assigned to Alpha and Bravo pods. Alpha pod was always opened 24 hours, while Bravo pod always closed at night. The emergency severity index (ESI) ranges from 1 to 5, and a lower ESI represents a more severe patient. Standard deviations are reported in parentheses.

Table 1.A.3: Peer Effects on Log Length of Stay

	(1)	(2)	(3)	(4)	(5)	(6)
Peer effect	0.115*** (0.035)	0.087* (0.049)	0.106*** (0.032)			
Peer effect, faster than median				0.142*** (0.047)	0.079 (0.069)	
Peer effect, slower than median				0.039 (0.098)	0.106 (0.124)	
Peer effect × self-managed			-0.053 (0.053)			
Peer effect, faster than median × self-managed						0.109* (0.062)
Peer effect, slower than median × self-managed						-0.081 (0.115)
Peer effect, faster than median × nurse-managed						0.128*** (0.045)
Peer effect, slower than median × nurse-managed						0.048 (0.093)
Self-managed			-0.081*** (0.029)			-0.061** (0.024)
Sample	Nurse-managed	Self-managed	Pooled	Nurse-managed	Self-managed	Pooled
Number of observations	129,274	129,798	259,072	129,274	129,798	259,072
Adjusted <i>R</i> -squared	0.485	0.378	0.404	0.485	0.378	0.404
Sample log length of stay mean	1.006	1.263	1.135	1.006	1.263	1.135

Note: I estimate physician peer effects depending on the organizational structure. I first estimate fixed effects in a regression of productivity on physician identities, adjusting for all possible interactions with resident, nurse, peer, and pod location. I then use these fixed effects for observations in the mechanical system where there is another physician in the same pod present to estimate a regression of log length of stay with peer effects. Models (1) and (4) are estimated for observations from nurse-managed teams, models (2) and (5) are estimated for observations from self-managed teams, and models (3) and (6) estimated from pooled observations. Models (4) to (6) split peer effects for peers that are slower or faster than median. All regressions in both stages are adjusted for hour of the day, day of the week, month-year dummies, Elixhauser score dummies, emergency, ED arrival volume, patient demographics, and patient triage time. All models are clustered for physician-resident-nurse trio identities.

Table 1.A.4: Log Length of Stay and on Log Time to First Order For Patients Arriving Near End of Shift

Outcome	(1)	(2)
	Log Length of Stay	Log Time to First Order
Near end of shift	-0.380*** (0.017)	-0.426*** (0.023)
Near end of shift × self-managed	0.415*** (0.018)	0.403*** (0.025)
Self-managed	-0.124*** (0.019)	0.011 (0.025)
Number of observations	307,554	293,839
Adjusted <i>R</i> -squared	0.389	0.148
Sample outcome mean	1.092	-0.625

Note: This table shows the effect on patients arriving near the end of shift for log length of stay and log time to first order. “Near end of shift” is defined as being within 2 hours of the end of shift for the physician assigned the patient in the nurse-managed system and for any of the two physicians who could choose the patient in the self-managed system. All regressions in both stages are adjusted for hour of the day, day of the week, month-year dummies, Elixhauser score dummies, emergency severity index, ED arrival volume, patient demographics, time patient spent in triage, and provider (physician-resident-nurse) identities. All models have fixed effects for and are clustered by physician identities.

Table 1.A.5: Other Outcomes For Patients Arriving Near End of Shift

Outcome	(1) 30-day mortality	(2) 14-day bounce- backs	(3) Hospital admissions	(4) Relative Value Units (RVUs)	(5) Log total costs
Near end of shift	0.0010 (0.0027)	-0.0038 (0.0060)	-0.0053 (0.0078)	-0.027 (0.030)	0.023 (0.026)
Near end of shift × self-managed	-0.0010 (0.0030)	0.0109 (0.0067)	0.0085 (0.0086)	0.070** (0.033)	0.049* (0.029)
Self-managed	0.0021 (0.0031)	-0.0134* (0.0069)	-0.0025 (0.0088)	-0.034 (0.029)	-0.029 (0.030)
Number of observations	308,616	308,616	308,616	248,016	277,974
Adjusted <i>R</i> -squared	0.302	-0.028	0.462	0.404	0.529
Sample outcome mean	0.019	0.067	0.272	2.698	6.721
Sample outcome standard deviation	0.138	0.250	0.445	1.323	1.515

Note: This table shows the effect on patients arriving near the end of shift for other outcomes. “Near end of shift” is defined as being within 2 hours of the end of shift for the physician assigned the patient in the nurse-managed system and for any of the two physicians who could choose the patient in the self-managed system. All regressions in both stages are adjusted for hour of the day, day of the week, month-year dummies, Elixhauser score dummies, emergency severity index, ED arrival volume, patient demographics, time patient spent in triage, and provider (physician-resident-nurse) identities. All models have fixed effects for and are clustered by physician identities.

1.B Theory Appendix

1.B.1 Model Setup

In this subsection, I outline an inclusive setup of the model that spans several assumptions, in both the nurse-managed system and the self-managed system. For detailed comments of the setup for the baseline model, refer to Section 1.3. Consider the following simple game, involving two physicians $j \in \{1, 2\}$ who work in a single pod at the same time:

1. The triage nurse (in the nurse-managed system) or physicians (in the self-managed system) commit to an assignment policy function if they are able to.
2. At time $t = 0$ both physicians receive one patient each, discovering $\theta_j \in \{\underline{\theta}, \bar{\theta}\}$, where $\bar{\theta} > \underline{\theta} > 0$. Type $\underline{\theta}$ occurs with probability p , and type $\bar{\theta}$ occur with probability $1 - p$.⁴⁷
3. Physicians commit to how long they will keep their initial patients (t_j).
4. Physicians send a message $m_j \in \{\underline{\theta}, \bar{\theta}\}$ to the triage nurse (in the nurse-managed system) or to each other (in the self-managed system) if they are able to report their types. Otherwise they send $m_j = \emptyset$.
5. With probability $\psi > 0$, physicians observe each other's θ_j . Physician types are never observed by the triage nurse.
6. Exactly one patient will arrive at time t_a distributed with uniform probability across $[\underline{\theta}, \bar{\theta}]$. The physician who receives this third patient, denoted by $\mathcal{J}(3)$, is determined as follows:
 - (a) In the nurse-managed system, the triage nurse assigns the patient based on physician censuses c_1 and c_2 (the number of patients they have, either 0 or 1, which is public information) at $t = t_a$. If she committed to an assignment policy function in Stage 1, she uses this. Otherwise, she decides assignment at t_a . If physicians can report their types, then her policy function can also use m_1 and m_2 .
 - (b) In the self-managed system, the physicians determine assignment.
 - i. If physicians cannot commit to an assignment policy function, each physician independently decides whether or not to choose the new at each time $t \geq t_a$.
 - ii. If physicians can commit to an assignment policy function, then the new patient is assigned according to this policy function at $t = t_a$. The policy function uses censuses c_1 and c_2 , and observed types $o_j \in \{\theta_j, \emptyset\}$ ($o_j = \theta_j$ with probability ψ , $o_j = \emptyset$ with probability $1 - \psi$). If physicians can report their types, then the policy function also uses m_j .

⁴⁷In this appendix, I will refer to θ_j as the type of physician j , since physicians start with one patient each and are otherwise identical.

7. Physicians complete their work on the one or two patients under their care and end their shifts. They receive the payoff

$$u_j^P = -(t_j - \theta_j)^2 - K_P(\theta_j) \mathbb{I}\{\mathcal{J}(3) = j\}, \quad (1.B.1)$$

where t_j is the time that physician j keeps his initial patient, $\theta_j \in \{\underline{\theta}, \bar{\theta}\}$ is the workload entailed by his patient and unobservable by the triage nurse, and $K_P(\theta_j) > 0$ is the cost of getting a potential third patient conditional on initial workload θ_j . I denote $K_P(\underline{\theta}) = \underline{K}_P$ and $K_P(\bar{\theta}) = \bar{K}_P$, and I impose that $\bar{K}_P > \underline{K}_P > 0$.

1.B.2 Nurse-managed System

In the nurse-managed system, the triage nurse assigns the new patient to a physician. I flexibly specify the triage nurse's utility as

$$u^N = -D \sum_{j \in \{1,2\}} (t_j - \theta_j)^2 - K_N(\theta_{j(3)}). \quad (1.B.2)$$

Notice the similarity between this utility function and that of the physicians, shown in (1.B.1). D is an indicator that allows the triage nurse to care about the treatment times of the first two patients as outcomes (if $D = 1$). Remember that the socially optimal discharge times for patients is $t_j = \theta_j$, which is universally agreed upon. The second term, $K_N(\theta)$, is the cost of assigning the new patient to a physician of type θ . I specify $K_N(\underline{\theta}) = 0$ and $K_N(\bar{\theta}) = \bar{K}_N$, and I impose $\bar{K}_N > 0$ but do not restrict the value of \bar{K}_N relative to $\bar{K}_P - \underline{K}_P$. This reflects that the triage nurse would like to assign the new patient to a physician with low workload.

The triage nurse's action is defined by an assignment policy function $\pi(c_1, c_2)$, where censuses $c_j \in \{0, 1\}$ are the numbers of patients the physicians have at time t_a .⁴⁸ To simplify notation, I impose that $\pi(0, 0) = \pi(1, 1) = \frac{1}{2}$ and $\pi \equiv \pi(0, 1) = 1 - \pi(1, 0)$. That is, when both physicians have equal censuses, the triage nurse should have no preference to send the new patient to one physician or the other, since she has no other information about who is less busy at that time. Also, probabilities must sum to 1. To be clear, $\pi = 1$ represents *ex post* efficiency in that if one physician has no patients and the other has one, the former physician is known to be less busy with certainty.

In what follows, I will first show that the analysis is simplified by the fact that each physician's best response function is unaffected by what he expects the other physician to do. I then consider three different scenarios of whether physicians can report their types and of whether the triage nurse can commit to an assignment policy function. In my preferred model, I assume that physicians cannot report their types but that the triage nurse can credibly commit to an assignment policy. Although θ_j is a single index known with certainty in this model, in practice it is difficult to verbalize

⁴⁸Below, I consider the scenario in which physicians can report their types to the triage nurse. In this case, the policy function takes the form $\pi(m_1, m_2)$, where $m_j \in \{\underline{\theta}, \bar{\theta}\}$.

or predict, let alone precisely report from physicians to the triage nurse at the outset of receiving each patient.⁴⁹ On the other hand, what the triage nurse does when she observes c_1 and c_2 is easily observable; the ED management may even set guidelines which instruct her to assign not entirely by censuses, and more importantly, beliefs about her behavior can easily be updated in practice by physicians when there is more than one new patient arriving during their shifts. Finally, I slightly modify the model to communicate the intuition for identification strategy of using the expected flow of future patients as a driver of foot-dragging.

1.B.2.1 Irrelevance of Peer Strategy or Type

One advantage of this simple two-type model is that the physician's strategy does not depend on the strategy or type of his peer. To see this, consider some assignment function $\pi(c_1, c_2)$, which we will assume for now is based on censuses. Again, this assignment function can be summarized by a single parameter $\pi \equiv \pi(0, 1)$. Any reasonable assignment rule sends the new patient to a less-busy physician, if there is one, with greater probability (i.e., $\pi > \frac{1}{2}$). It is easy to see that a high-type physician will then discharge his patient at $\bar{t} = \bar{\theta}$, because discharging his patient earlier only increases his chances of getting the new patient and discharging later has no benefit either.

Lemma 1. *A high-type physician always discharges his patient at $\bar{t} = \bar{\theta}$.*

The next step is to solve for the best response function of the index physician who is of low type, given his peer's strategy. Denote the discharge time that a low-type peer will choose as $\underline{t}_{-j} \in [\underline{\theta}, \bar{\theta}]$. The expected utility of the index physician is

$$\mathbb{E}[u_j^P(t_j; \pi, \underline{\theta})] = -(t_j - \underline{\theta})^2 - \underline{K}_P \Pr\{\mathcal{J}(3) = j | t_j, \pi\},$$

The probability that $\mathcal{J}(3) = j$ can be written more explicitly as

$$\Pr\{\mathcal{J}(3) = j | t_j, \pi\} = \begin{cases} \frac{1}{2} \left(\frac{t_j - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + \pi \left(\frac{\underline{t}_{-j} - t_j}{\bar{\theta} - \underline{\theta}} \right) + \left(\frac{1}{2}p + \pi(1-p) \right) \left(\frac{\bar{\theta} - t_j}{\bar{\theta} - \underline{\theta}} \right), & t_j < \underline{t}_{-j} \\ \frac{1}{2} \left(\frac{\underline{t}_{-j} - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + ((1-\pi)p + \frac{1}{2}(1-p)) \left(\frac{t_j - \underline{t}_{-j}}{\bar{\theta} - \underline{\theta}} \right) + \left(\frac{1}{2}p + \pi(1-p) \right) \left(\frac{\bar{\theta} - t_j}{\bar{\theta} - \underline{\theta}} \right), & t_j > \underline{t}_{-j} \end{cases}, \quad (1.B.3)$$

which is continuous at $t_j = \underline{t}_{-j}$. This expression represents flow probabilities divided among three potential windows of time and uses the fact that the peer is low-type with probability p .

Both \underline{t}_{-j} and p additively affect expected utility, but they do not affect the first-order condition

⁴⁹I also assume that there is no credible way for physicians to report each other's workloads to the triage nurse. Moore and Repullo (1988) have formalized such a subgame perfect mechanism. However, with either limited financial incentives or social incentives such as reciprocity, such a mechanism may not be implementable.

with respect to \underline{t}_j . In particular,

$$\frac{\partial}{\partial \underline{t}_j} \Pr \{ \mathcal{J}(3) = j \mid \underline{t}_j, \pi \} = \frac{1}{\bar{\theta} - \underline{\theta}} \left(\pi - \frac{1}{2} \right). \quad (1.B.4)$$

The flow probability of receiving the third patient is always reduced by $\pi - \frac{1}{2}$ by increasing \underline{t}_j . This implies that his best response does not depend on the action of his (low-type) peer, nor on the probability that his peer is low-type. Consequently, it also does not matter whether he knows his peer's type with certainty, and so the sequence in which physicians first choose \underline{t}_j and then observe θ_{-j} with probability ψ is immaterial.

Lemma 2. *The best response of a low-type physician in the nurse-managed system does not depend on the action of his peer nor the probability that his peer is low-type.*

1.B.2.2 No Physician Reporting, No Triage Nurse Commitment

Without physician reporting or triage nurse commitment, in equilibrium, the triage nurse chooses the optimal assignment policy $\pi(c_1, c_2)$ at time t_a , given physician discharge strategies \underline{t}^* and \bar{t}^* for low- and high-type physicians, respectively, and given censuses c_1 and c_2 . Given this choice, summarized as $\pi^* = \pi^*(0, 1)$, physicians choose the optimal discharge strategies \underline{t}^* and \bar{t}^* .

Proposition 3. *In the Perfect Bayesian Equilibrium in the nurse-managed system with no physician reporting and no triage nurse commitment, the triage nurse always assigns the new patient to the physician with census 0 when there is another physician with census 1, i.e., $\pi^* = 1$. Low-type physicians will foot-drag, choosing $\underline{t}^* > \underline{\theta}$, as in Equation (1.B.5).*

The triage nurse's assignment policy is simple: She will assign the new patient to the physician with no patients if the other has one patient (i.e., $\pi = 1$). In this case, given any \underline{t}^* and \bar{t}^* , she expects that the physician with $c_j = 0$ must be low-type whereas the one with $c_{-j} = 1$ must be high type. Otherwise, if $c_1 = c_2$, she cannot distinguish the two physicians and will assign the patient with equal probability to each.

In equilibrium, physicians will discharge their initial patients with this knowledge. As stated in Lemma 1, high-type physicians will never want to mimic low-type physicians and will discharge their patients at time $t = \bar{\theta}$. On the other hand, low-type physicians will want to mimic high-type physicians at least temporarily by keeping their patients longer than socially optimal, since this reduces his likelihood of getting the new patient. (Low-type) physicians do not consider the type of their peer because of Lemma 2. The first-order condition of a low-type physician's problem of $\max_{\underline{t}_j} \mathbb{E} [u_j^P(t_j; \underline{\theta})]$ yields

$$\underline{t}^* = \underline{\theta} + \frac{K_P}{4(\bar{\theta} - \underline{\theta})}. \quad (1.B.5)$$

This reflects the fact that there is always a first-order gain to reducing the likelihood of getting the new patient, compared to a second-order loss to prolonging the patient stay to more than socially

optimal. That is, there will always be foot-dragging. In fact, for sufficiently high \underline{K}_P or sufficiently low $\bar{\theta} - \underline{\theta}$, there may be full pooling in that all physicians will discharge their patient at $t = \bar{\theta}$.

In this subgame perfect equilibrium, the triage nurse correctly assigns patients to physicians who are less busy when she observes $c_1 \neq c_2$. But by doing so, she incentivizes physicians to foot-drag and actually reduces the probability of seeing $c_1 \neq c_2$ when $\theta_1 \neq \theta_2$. That is, while her assignment is *ex post* efficient, it is *ex ante* inefficient.

1.B.2.3 No Physician Reporting, Triage Nurse Commitment

I now allow for the triage nurse to commit to a policy function π . The equilibrium in this case is the same as in the previous case without commitment, except that the triage nurse chooses π_C^* at $t = 0$ and not at $t = t_a$.

Proposition 4. *In the Perfect Bayesian Equilibrium in the nurse-managed system with no physician reporting but triage nurse commitment, the triage nurse assigns the new patient to the physician with census 0 when there is another physician with census 1 with some probability π_C^* as given by Equation (1.B.8), which lies between $\frac{1}{2}$ and 1. A low-type physician foot-drag weakly less than under no triage nurse commitment, discharging his patient at \underline{t}^* that is earlier and closer to $\underline{\theta}$.*

To analyze this case, first note that the triage nurse will still never want to send the new patient with greater probability to a physician with $c_j > c_{-j}$. It therefore still holds that high-type physicians will never want to mimic low-type physicians, and that low-type physicians have some reason to mimic high-type ones. To see this, for any given π_C , the first-order condition for a low-type physician yields

$$\underline{t}^* = \underline{\theta} + \frac{\underline{K}_P}{2(\bar{\theta} - \underline{\theta})} \left(\pi_C - \frac{1}{2} \right). \quad (1.B.6)$$

The first-order gain in temporary mimicry still exists relative to the second-order loss, as long as $\pi > \frac{1}{2}$.

When the triage nurse commits to an assignment policy function, she will choose π such that her expected utility is maximized. Her expected utility is

$$\begin{aligned} \mathbb{E} [u^N(\pi_C)] &= -Dp(\underline{t}^* - \underline{\theta})^2 - (1-p)^2 \bar{K}_N - \\ &\quad 2p(1-p) \bar{K}_N \left[\frac{1}{2} \left(\frac{\underline{t}^* - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + (1 - \pi_C) \left(\frac{\bar{\theta} - \underline{t}^*}{\bar{\theta} - \underline{\theta}} \right) \right]. \end{aligned} \quad (1.B.7)$$

Substituting (1.B.6) into (1.B.7) and solving the first-order condition yields the optimal assignment rule under commitment

$$\pi_C^* = \frac{1}{2} + \frac{4p(1-p) \bar{K}_N (\bar{\theta} - \underline{\theta})^2}{\underline{K}_P (Dp \underline{K}_P + 4p(1-p) \bar{K}_N)}. \quad (1.B.8)$$

It is easy to see that π_C^* given by (1.B.8) can be less than 1, which implies that the triage nurse sometimes sends the new patient to the physician with census 0 even when there is another one

with census 1. With $\pi_C^* < 1$, by (1.B.6), \underline{t}^* is lower than when there is no triage nurse commitment.

The important general point is that for some parameters the triage nurse will commit to an assignment policy function $\pi_C^* < 1$. Even if she only cares about the assignment of the third patient, as in Equation (1.B.9), she will commit to an *ex post* inefficient assignment policy in order to improve *ex ante* assignment. Under triage nurse commitment to $\pi_C^* < 1$, the degree of foot-dragging by low-type physicians, in Equation (1.B.6), will be lower than under no commitment, stated in Equation (1.B.5). The intuition for this is similar to Milgrom and Roberts' (1988) finding that managers will sometimes choose to ignore valuable but distortable information from "influence activities."

This result is even stronger when the triage nurse cares about length of stay for the initial patients (i.e., $D = 1$). To see this, consider the optimal policy function if the triage nurse does not care about lengths of stay for the first two patients as outcomes and only cares about the assignment of the third patient (i.e., $D = 0$):

$$\pi_{C|D=0}^* = \frac{1}{2} + \frac{(\bar{\theta} - \underline{\theta})^2}{K_P}. \quad (1.B.9)$$

The triage nurse's choice of $\pi_{C|D=0}^*$ only depends on the *low-type physician's* cost of getting the new patient, because it is this physician that will engage in foot-dragging and distort information. In the case with $D = 1$, in (1.B.8), $\pi_{C|D=1}^*$ also increases as \bar{K}_N or as p is closer to $\frac{1}{2}$. Since she cares about foot-dragging as an outcome, she will increase $\pi_{C|D=1}^*$ as she cares more about inefficient assignment (as \bar{K}_N is higher) or as it is more likely she will encounter two physicians with different censuses (p is closer to $\frac{1}{2}$). Importantly, comparing (1.B.9) and (1.B.8), $\pi_{C|D=0}^* > \pi_{C|D=1}^*$. Also caring about foot-dragging as an outcome lowers her choice of π_C^* to reduce foot-dragging further.

1.B.2.4 Physician Reporting, Triage Nurse Commitment

I next consider the case in which physicians can also report their types as $m_j \in \{\underline{\theta}, \bar{\theta}\}$, which is a mechanism design problem without transfers. In the equilibrium in this case, the triage nurse provides a menu $\{t(\underline{\theta}), t(\bar{\theta}); \pi(m_1, m_2)\}$ to physicians, subject to an incentive compatibility constraint for physicians to report $m_j = \theta_j$ truthfully. Because the triage nurse immediately distributes the new patient upon arrival, there is an additional "intertemporal" constraint in that a physician who lies and then later reveals that he is lying can still be punished sufficiently so that he will not want to engage in this strategy. Given this menu, physicians will tell the truth.⁵⁰

Proposition 5. *In the Perfect Bayesian Equilibrium in the nurse-managed system with physician reporting and triage nurse commitment, the triage nurse implements truth-telling by setting the assignment policy π_R^* equal to the assignment policy implied by intertemporal incentive compatibility, π_{IT}^* , given in Equation (1.B.13). Given truth-telling and discrete types, there is no foot-dragging,*

⁵⁰ Again, expectations over the peer's type are conveniently ignored due to Lemma 2.

i.e., $\underline{t}^* = \underline{\theta}$. Assignment is still ex post inefficient with $\pi_R^* < 1$, but ex ante assignment efficiency is improved relative to no physician reporting.

By Lemma 1, there is no incentive compatibility constraint for high-type physicians, but there is one for low-type physicians. Specifically, the utility for a low-type physician under truth-telling must be at least as great as his utility if he were to report that he is a high-type physician, where I again use the notation $\pi \equiv \pi(\underline{\theta}, \bar{\theta})$ to summarize any assignment policy:

$$-(t(\bar{\theta}) - \underline{\theta})^2 - \frac{1}{2}\underline{K}_P \leq -\pi_{IC}\underline{K}_P. \quad (1.B.10)$$

Using the fact that $t(\bar{\theta}) = \bar{\theta}$ by Lemma 1,⁵¹ this incentive compatibility constraint is actually binding at the optimal triage nurse assignment when she only cares about foot-dragging as a signal, stated in Equation (1.B.9):

$$\pi_{IC}^* = \frac{1}{2} + \frac{(\bar{\theta} - \underline{\theta})^2}{\underline{K}_P}. \quad (1.B.11)$$

Note that although π_{IC}^* equals $\pi_{C|D=0}^*$ in Equation (1.B.9), the intuitions for the two expressions are different. In the case of π_{IC}^* , the constraint sets utility to be the same for a low-type physician under truth-telling and under reporting to be high-type. In the case of $\pi_{C|D=0}^*$, triage nurse utility happens to be maximized at an assignment policy that causes a low-type physician to foot-drag midway with $\underline{t}^* = \underline{\theta} + (\bar{\theta} - \underline{\theta})/2$. Also recall that $\pi_{C|D=0}^* > \pi_{C|D=1}^*$; so $\pi_{IC}^* \leq \pi_C^*$.

However, there is a second “intertemporal” incentive compatibility constraint that results from the facts that reports of θ_j and discharge times t_j are intertemporally separated and that the triage nurse is limited in her ability to punish a physician who reports $\theta_j = \bar{\theta}$ but discharges his patients before $t(\bar{\theta}) = \bar{\theta}$. That is, a low-type physician may report that he is high-type but discharge his patient at $\hat{t} < \bar{\theta}$. Without punishment (i.e., if he simply reverts to having the truth-telling flow probability of $\pi(\underline{\theta}, \bar{\theta})$ after t) he would be strictly better off by this scheme. To prevent this, there needs to be punishment, but the highest flow probability that the triage nurse can use from that point on is 1.

The second constraint is thus

$$-(\hat{t} - \underline{\theta})^2 - \underline{K}_P \left[\frac{1}{2} \left(\frac{\hat{t} - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + \frac{\bar{\theta} - \hat{t}}{\bar{\theta} - \underline{\theta}} \right] \leq -\pi \underline{K}_P. \quad (1.B.12)$$

This can be addressed by first solving for the optimal “cheating” \hat{t}^* under full punishment:

$$\hat{t}^* = \underline{\theta} + \frac{\underline{K}_P}{4(\bar{\theta} - \underline{\theta})},$$

⁵¹In some mechanism design problems, this value of a high-type agent could be distorted upwards. However, because I have assumed that there is 0 probability that the new patient will arrive after $t = \bar{\theta}$, this mechanism cannot be enforced.

which of course is the same as the optimal discharge time with no physician reporting or triage nurse commitment, stated in (1.B.5). This time can then be substituted into (1.B.12) in order to state the constraint on π :

$$\pi_{IT}^* = 1 - \frac{\underline{K}_P}{16(\bar{\theta} - \underline{\theta})^2}. \quad (1.B.13)$$

Note that $\pi_{IT}^* < 1$ for $\underline{K}_P > 0$. The intuition for this is that, with intertemporal incentive compatibility, the triage nurse can never implement $\pi = 1$, because if she did, then a low-type physician will always be better off by reporting to be high-type for some time. Revealing that he lied entails “punishment” which cannot be worse than the $\pi = 1$ he would have gotten with truth-telling anyway.⁵²

It can be shown that $\pi_{IC}^* \leq \pi_{IT}^*$. In particular, for $\underline{K}^P / (\bar{\theta} - \underline{\theta})^2 \in (0, 4)$, when there is some temptation to foot-drag but when $\underline{t}^* < \bar{\theta}$ in the case without physician reporting or triage nurse commitment, as in Section 1.B.2.3 and Equation (1.B.5), $\pi_{IC}^* < \pi_{IT}^*$. The intuition for this is that it is weakly more difficult to implement the intertemporal constraint than the standard incentive compatibility constraint because it is always at least as easy for low-type physicians to “temporarily” lie rather than “fully” lie. Thus, the intertemporal constraint will always be binding.

The assignment policy under physician reporting is therefore $\pi_R^* = \min(\pi_{IC}^*, \pi_{IT}^*) = \pi_{IT}^*$. Again, the assignment policy under no physician reporting but triage nurse commitment is equivalent to the assignment policy implied by the constraint with “full” lying: $\pi_{C|D=0}^* = \pi_{IC}^*$. At first glance, this suggests that assignment is less efficient *ex post* under physician reporting than under no reporting: $\pi_{C|D=0}^* < \pi_R^*$. This may not hold if the triage nurse also cares about foot-dragging on the initial patients ($D = 1$) under no reporting, since $\pi_{C|D=0}^* > \pi_{C|D=1}^*$.

However, the more important point is that under reporting and truth-telling, *ex post* assignment efficiency is also *ex ante* assignment efficiency, since physicians do not foot-drag. Further, it can be shown that the triage nurse’s *ex ante* utility is strictly greater with physician reporting than with no reporting. In order to formalize this, I consider $\mathbb{E}[u^N]$, where u^N is given in Equation (1.B.2), which simplifies to $\bar{K}^N \Pr\{\theta_{j(3)} = \bar{\theta}\}$ when $D = 0$. Figure 1.B.2 shows the difference in this value between physician reporting and no reporting, normalizing $\bar{K}^N = 1$. Note that this efficiency gain is strictly greater when $D = 1$, because the triage nurse also cares about foot-dragging on the initial patients as negative outcomes, of which there is none under reporting.

In summary, the result in Proposition 5 derives from the fact that the triage nurse has access to better information from physician reports about their types. Subject to maintaining truth-telling, specifically truth-telling that is robust to “partial” or “temporary” lying, she can implement an *ex ante* assignment policy that is more efficient than the one without physician reporting, when she must balance *ex post* assignment efficiency with distortion of signals by physicians. In both cases,

⁵²In contrast, $\pi_{IC}^* = 1$ for $\underline{K}^P / (\bar{\theta} - \underline{\theta})^2 \leq 2$. In this parameter space, a low-type physician is better off by telling the truth than by “fully” lying by reporting that he is high-type and keeping his patient until $t(\bar{\theta}) = \bar{\theta}$. Similarly, $\pi_C^* = 1$ for $\underline{K}^P / (\bar{\theta} - \underline{\theta})^2 \leq 2$, because the temptation to foot-drag is sufficiently low so that the triage nurse is better off by incurring the maximum foot-dragging and the highest *ex post* assignment efficiency.

she is limited by the ability of physicians to distort signals or misreport the truth. In this sense, there is also a parallel intuition between Milgrom and Roberts' (1988) prediction and the standard mechanism design feature of information rents.

In this two-type model, there is no foot-dragging because the triage nurse uses π_R^* in order to implement truth-telling. There is no point in using $t(\underline{\theta})$ to implement truth-telling, because doing so would only make a low-type physician worse off under truth-telling, and $t(\bar{\theta})$ is irrelevant because of the intertemporal incentive compatibility constraint. However, I will show in Section 1.B.4 that this does not hold for continuous types, because with local incentive compatibility constraints, the benefit of foot-dragging ("full" lying in the mechanism design framework) at the truth ($t_j = \theta_j$) is first-order while its cost is second order.

1.B.2.5 Flow of Expected Future Work

This simple model assumes a single patient will arrive in the interval $t \in [\underline{\theta}, \bar{\theta}]$, which is convenient for capturing the temptation for moral hazard by low-type physicians. However, there are of course in practice usually many more new patients, and my main identification for foot-dragging will be the response of lengths of stay to the flow of expected future work. One way to capture the intuition of expected future work is to modify the model so that a new patient is expected to arrive in the interval $t \in [\underline{\theta}, \underline{\theta} + \Delta t]$, maintaining the assumption of a single new patient. I also allow Δt to be greater or less than $\bar{\theta} - \underline{\theta}$, although I assume that physicians may only be assigned the new patient prior to $t = \bar{\theta}$ for $\Delta t > \bar{\theta} - \underline{\theta}$ and focus on interior solutions for $\Delta t < \bar{\theta} - \underline{\theta}$.⁵³

I again focus on the behavior of low-type physicians, who has an incentive to foot-drag and mimic high-type physicians. It is easy to see that all denominators in fractions in Equation (1.B.3) should now be Δt instead of $\bar{\theta} - \underline{\theta}$, and $\bar{\theta}$ should be replaced by $\underline{\theta} + \Delta t$. Equation (1.B.4) should then have Δt in the denominator rather than $\bar{\theta} - \underline{\theta}$, at least for interior $\underline{t}_j \in [\underline{\theta}, \bar{\theta} + \Delta t]$:

$$\frac{\partial}{\partial \underline{t}_j} \Pr \{j(3) = j \mid \underline{t}_j \in [\underline{\theta}, \bar{\theta} + \Delta t], \pi\} = \frac{1}{\Delta t} \left(\pi - \frac{1}{2} \right).$$

Respective denominators in the foot-dragging Equations (1.B.5) and (1.B.6) should also now have Δt instead of $\bar{\theta} - \underline{\theta}$ for interior solutions $\underline{t}^* \in [\underline{\theta}, \bar{\theta} + \Delta t]$. For example, in the baseline scenario of no physician reporting but triage nurse commitment, the equilibrium foot-dragging by low-type physicians is

$$\underline{t}^* = \underline{\theta} + \frac{K_P}{2\Delta t} \left(\pi - \frac{1}{2} \right).$$

This slight modification communicates the intuition that as the expected future work increases (decreases), the marginal temptation to foot-drag increases (decreases) because the certainty of

⁵³I maintain the assumption of a single new patient, so that I do not have to consider capacity constraints and physician strategy for subsequent patients, in order to keep the model simple. I also focus on interior solutions so that this single patient is relevant for comparative statics.

receiving a new patient within each infinitesimal unit of time around discharge is greater (smaller).

1.B.3 Self-managed System

In this subsection, I will analyze foot-dragging and patient assignment in the self-managed system. I assume the same physician utilities and information structure (i.e., that they observe each other's θ_j with probability $\psi > 0$) as I did for the nurse-managed system. The only difference will be that the two physicians, not a triage nurse, are responsible for deciding who gets the new patient. In what follows, I will also consider analagous cases in which physicians may or may not report θ_j to each other and in which they may or may not be able to commit to an assignment policy function based on censuses. The key difference between results for all of these cases and corresponding results for the nurse-managed system derives from physicians *both* observing each other's θ_j with probability $\psi > 0$ *and* being able to use that information in patient assignment.

The assignment of patients by the physicians themselves deserves further mention. In the case in which physicians are unable to commit to an assignment policy, I represent assignment as a non-cooperative bargaining game in which physicians can choose to see the new patient. At any time after the new patient has arrived, as long as the patient remains unchosen, either physician may choose to see the new patient or wait. If one physician chooses the patient, he gets the patient with probability 1. If they both choose the patient at the same time, they each get the patient with probability $\frac{1}{2}$. This game is very much motivated by Rubinstein's (1982) non-cooperative bargaining game in which two players with complete information about each other's costs bargain over a good that declines in value over time. Aided by a setup that is simpler than Rubinstein's, I will also extend its analysis to consider incomplete information about peer types.

In the other case in which physicians can commit to an assignment policy, I represent an assignment policy function that is quite similar to the one I defined previously for the nurse-managed system. Here the probability of assignment to physician 1 is a function of censuses and potential observations of type $o_j \in \{\emptyset, \theta_j\}$, where recall that both types are observed with probability ψ . That is, the policy function takes the form $\pi(c_1, c_2, o_1, o_2)$. It is easy to see that physicians should commit to $\pi(c_1, c_2, \underline{\theta}, \bar{\theta}) = 1$ for all c_1 and c_2 . Similar to before, the one relevant parameter to be solved for is $\pi \equiv \pi(0, 1, \emptyset, \emptyset)$. Commitment in this case involves physicians jointly determining an assignment rule to maximize expected utility prior to knowing their types.

1.B.3.1 No Physician Reporting, No Policy Commitment

I first consider the case in which physicians can neither report their types nor commit to an assignment policy. In order to microfound the assignment process, I need to elaborate on the cost of waiting to treat a patient who has arrived. Denote τ as the time that has elapsed since the new patient arrived. Assume that both physicians incur a cost of having the new patient remain untreated at each time τ , increasing over τ , and also assume now that the cost of getting the new patient also increases over τ but not as quickly.

Extended Physician Utility

Formally, I extend physician utility as

$$u_j^P = -(t_j - \theta_j)^2 - W_j(\theta_j; \tau^*, j(3)) - K_P(\theta_j) \mathbb{I}\{j(3) = j\}.$$

The new second term $W_j(\cdot)$ is a generalized cost of waiting. Denoting τ^* as the elapsed time that it took for the new patient to be chosen by someone, and $j(3)$ as the physician who chose the patient, W_j is simply integrated over each τ .

$$W_j(\theta_j; \tau^*, j(3)) = \int_0^{\tau^*} \omega_j(\tau; \theta_j, \tau^*, j(3)) d\tau,$$

$\omega_j(\tau; \theta_j, \tau^*, j(3))$ is a flow cost that depends on τ and whether physician j received the patient at τ or not:

$$\omega_j(\tau; \theta_j, \tau^*, j(3)) = \begin{cases} w\tau, & \text{if } \tau \neq \tau^* \\ k_0(\theta_j) + k\tau, & \text{if } \tau = \tau^*, j = j(3) \\ 0, & \text{if } \tau = \tau^*, j \neq j(3). \end{cases}$$

$w\tau$ is the flow cost of waiting imposed on both physicians as long as the patient remains unchosen. $k_0(\theta_j) + k\tau$ is the initial flow cost of seeing the patient, where I denote $\underline{k}_0 = k_0(\underline{\theta})$ and $\bar{k}_0 = k_0(\bar{\theta})$ for brevity and assume $\underline{k}_0 < \bar{k}_0$. Note that since $\tau = \tau^*$ with mass of 0, I can simplify the cost-of-waiting integral to $W_j(\tau^*) = \frac{1}{2}w\tau^{*2}$.

I assume that the flow costs of seeing that patient, starting with the initial cost $k_0(\theta_j)$ if the patient is seen immediately, integrate to $K_P(\theta_j)$. This ensures that this extended utility is consistent with its simpler form in Equation (1.B.1) in the nurse-managed system, where physicians are assigned patients immediately by the triage nurse at $\tau = 0$.⁵⁴ To ensure single crossing, I impose that k and the continued flow cost of seeing the new patient are both less than w . Finally I assume that physicians are responsible for taking care of patients up to $t = \bar{\theta} + \bar{k}_0/(w - k)$ to ensure that any patient arriving up to $t = \bar{\theta}$ could be chosen by a physician.

Non-cooperative Assignment by Physicians

In order to analyze assignment by physician choice, I consider four different cases. The first two cases occur when physicians observe each other's types, which happens with probability ψ . Case 1 is that physicians observe that one is type $\underline{\theta}$ and the other is type $\bar{\theta}$ and proceeds quite similarly to Rubinstein's (1982) setup with complete information. Physicians will infer that a low-type physician will weakly prefer choosing the patient at $\tau = \underline{k}_0/(w - k)$, while a high-type physician will weakly

⁵⁴I could also consider that the triage nurse might delay the assignment of patients to physicians if she can gain more information about their types. However, this would not add any useful intuition for the nurse-managed system, and of course, the setup of physicians starting with one patient each is itself an abstraction.

prefer choosing the patient at $\bar{\tau} = \bar{k}_0 / (w - k)$ and will strictly prefer waiting at $\underline{\tau} < \bar{\tau}$. Therefore, by the subgame where physicians choose to see the patient at $\underline{\tau}$, a low-type physician will be assigned the patient with probability 1. Given that there is a cost to waiting, a low-type physician will prefer to see the patient immediately than to wait until $\underline{\tau}$. So in equilibrium, a low-type physician will choose the patient at $\tau^* = 0$ with probability 1.

Case 2 is that physicians observe that they are both the same type $\theta \in \{\underline{\theta}, \bar{\theta}\}$. There exists no pure strategy Nash equilibrium here. To see this, if there exists a Nash equilibrium, under symmetry both physicians should choose the patient at some $\tau^*(\theta)$. But one physician would always be strictly better off by waiting and letting the other physician choose the patient. There does exist a mixed strategy Nash equilibrium though. I generally denote the probability with which each physician will choose the patient at each elapsed time τ as $q(\theta; \theta, \tau)$, where the first argument is the type θ_j of the index physician (in this case shared) and the second argument denotes the observation $\theta_{-j} \in \{\theta_{-j}, \emptyset\}$ of the peer's type (in this case observed). To satisfy the conditions for indifference between choosing and not choosing,

$$q(\theta; \theta, \tau) = \frac{\max\{0, w\tau - (k_0(\theta) + k\tau)\}}{w\tau - (k_0(\theta) + k\tau)/2}, \quad (1.B.14)$$

From Equation (1.B.14), note that $q(\theta, \tau) > 0$ only when the flow cost of waiting is inefficiently greater than the flow cost of initially treating the patient, i.e., $\tau > k_0(\theta) / (w - k)$. Also, although $q(\theta, \tau)$ increases with τ , $\lim_{\tau \rightarrow \infty} q(\theta) = (w - k) / (w - k/2) < 1$.

The next two cases occur in the subgame, occurring with probability $1 - \psi$, in which physicians do not observe each other's types. As in the nurse-managed system, a high-type physician will still discharge his patient at $t = \bar{\theta}$. Suppose that a low-type physician discharges his patient at $t = \underline{t}^*$. Types can then be perfectly deduced only during times $t \geq \underline{t}^*$. So prior to \underline{t}^* , we have a game of incomplete information, and if the new patient arrives at $t_a < \underline{t}^*$, physicians will have an incentive to wait because types are unknown and therefore certain times for patient choice (i.e., Case 1) are impossible. Thus, during any $t \in T_\emptyset[t_a, \underline{t}^*]$, physicians must decide whether to choose patients while peer types are unknown. Case 3 considers the choice equilibrium strategy for a low-type physician during this time. As in Case 2, he will purely wait if $\tau \leq \underline{k}_0 / (w - k)$ and will engage in a mixed strategy if $\tau > \underline{k}_0 / (w - k)$. Similar to Equation (1.B.14),

$$q(\underline{\theta}; \emptyset, \tau) = \min \left\{ 1, \frac{\max\{0, w\tau - (\underline{k}_0 + k\tau)\}}{p(w\tau - (\underline{k}_0 + k\tau)/2)} \right\}. \quad (1.B.15)$$

Equation (1.B.15) shows that a low-type physician is actually more likely to choose the new patient when he does not know that his peer's type relative to when he knows his peer is also low-type (but obviously less likely than when he knows his peer is high-type). Also, if the probability p of having a low-type peer is low enough (i.e., if $p < (w - k) / (w - k/2)$), he may even choose the patient with certainty at some point.

Case 4 considers the strategy for a high-type physician during the same $T_\emptyset = [t_a, \underline{t}^*]$. For a given elapsed time since arrival τ , note that both high-type and low-type physicians cannot be engaging in a mixed strategy. This would require

$$\frac{w\tau - (\underline{k}_0 + k\tau)}{w\tau - (\underline{k}_0 + k\tau)/2} = pq(\underline{\theta}; \emptyset, \tau) + (1-p)q(\bar{\theta}; \emptyset, \tau) = \frac{w\tau - (\bar{k}_0 + k\tau)}{w\tau - (\bar{k}_0 + k\tau)/2},$$

which is impossible for finite τ , given $\underline{k}_0 < \bar{k}_0$. So in equilibrium, high-type physicians will wait until after the time when low-type physicians should have chosen the new patient with certainty, which is implied in (1.B.15) to be

$$\underline{\tau}^* = \frac{(1 - \frac{1}{2}p) \underline{k}_0}{(1-p)w - (1 - \frac{1}{2}p)k}.$$

Note that types will then be revealed at $t_R = \min(\underline{t}^*, t_a + \underline{\tau}^*)$.

Finally, note that if the patient is still unchosen at t_R when physician types are revealed, then by definition, both physicians must be high type, and they will know this. They should mix with probability as defined in (1.B.14) as in Case 2. They could wait until some time after t_R until they start mixing, or they could immediately start mixing at t_R , depending on whether $t_R - t_a > \bar{k}_0/(w - k)$. As in Case 2, each physician will never choose the patient with certainty.

No Incentive for Foot-dragging

Considering all 4 cases above, it is clear that low-type physicians can never hope to have the new patient assigned to a high-type peer. This is true even when types are unknown by peers, because as shown in Case 4, high-type physicians will still wait until any low-type physician would have chosen the patient with certainty before choosing the patient with any positive probability.

Another point that is obvious from the above analysis of physician assignment without policy commitment is that physicians will often delay choosing patients, if they know that they are equally busy or if they are unsure how busy their peer is. In fact, under this delay, patients may sometimes never be chosen, given that there is no commitment to choose patients eventually. The delay in choosing patients represents the channel of “free-riding.” However, this represents an *ex ante* utility loss at the stage that low-type physicians decide \underline{t}^* , since physicians only start placing some positive probability on choosing the new patient when they would have chosen the patient anyway if there were no peer.

Proposition 6. *In the Perfect Bayesian Equilibrium in the self-managed system with no physician reporting or commitment to an assignment policy, there will be no foot-dragging, i.e., $\underline{t}^* = \underline{\theta}$. If the patient is chosen and if there is a low-type physician, assignment will always be to a low-type physician. However, physicians will wait to choose the new patient (free-ride) if they are of the same type.*

Low-type physicians have no incentive to conceal their types by foot-dragging for two reasons: First, low-type physicians can never hope to have a potential high-type peer choose the new patient before them. Second, concealing that they are low-type (foot-dragging) only leads to free-riding, which represents an *ex ante* utility loss. As in the nurse-managed system, assignment is completely *ex post* efficient in the sense that patients are never assigned to physicians with lower censuses when censuses differ. In addition, given that there is no foot-dragging, this also implies *ex ante* efficiency. However, there is a new “assignment” inefficiency of free-riding in that physicians may delay seeing patients and sometimes not even get to see them despite preferring to had there been no peer. In this model, free-riding only occurs when physicians are of the same type, since types can always be inferred by the time the new patient arrives at $t \in [\underline{\theta}, \bar{\theta}]$, as low-type physicians never foot-drag.⁵⁵

Remark on Continuous Types

I will conclude this subsection with a remark on continuous types in order to show that with continuous types the intuition for the stark result in Proposition 6 does not hold. To see this, first note that with continuous types, even if types are unknown, the probability that a physician has the same type as his peer has mass 0. Thus, physicians should choose patients with pure strategies in equilibrium.

When physician do not observe each other’s types (with probability $1 - \psi$), physicians then use peer signals to infer θ_{-j} , and by subgame perfection physician j will choose the new patient at $\tau = 0$ if and only if $c_j > c_{-j}$. We now have a similar situation as in the traditional system with no triage nurse commitment, in Section 1.B.2.3, where there is a first-order gain to foot-dragging. So in equilibrium, there should be no free-riding but positive foot-dragging. However, recall that with probability ψ , physicians observe each other’s types. In this case, the physician j will choose the new patient at $\tau = 0$ if and only if $\theta_j > \theta_{-j}$, regardless of c_j and c_{-j} . Because of this, physicians will foot-drag less than they would have under the traditional system with no triage nurse commitment.

1.B.3.2 No Physician Reporting, Policy Commitment

From the analysis in Section 1.B.3.1, it is clear that without physician commitment to an assignment policy, there could be large welfare losses in the form of free-riding and patients going untreated. This suggests scope for improvement by committing to an assignment policy. As a practical rationale, physicians often divide work before they can deduce each other’s true workloads if new patients need to be seen in a timely manner.

Introduced above, the policy function takes the form $\pi(c_1, c_2, o_1, o_2)$, where o_j is type of physician j if observed and null otherwise. The following are obvious: $\pi(c_1, c_2, \underline{\theta}, \bar{\theta}) = 1$ for all c_1 and

⁵⁵In practice, types may not be perfectly deduced (which would also happen in the model if the new patient could arrive at $t < \underline{\theta}$), and this could support free-riding. However, as shown in the main chapter, free-riding does not appear to be significant empirically, which suggests that physicians can commit to an assignment policy or have sufficient information about each other’s types (either by censuses or observations of true workload).

c_2 ; $\pi(c_1, c_2, \theta_1, \theta_2) = \frac{1}{2}$ for $\theta_1 = \theta_2$ and all c_1 and c_2 ; and $\pi(c_1, c_2, \emptyset, \emptyset) = \frac{1}{2}$ for $c_1 = c_2$.⁵⁶ As in the nurse-managed system, the assignment policy can then be represented by a single parameter $\pi_{C,Self} \equiv \pi(0, 1, \emptyset, \emptyset)$. In equilibrium, physicians choose the optimal assignment policy $\pi_{C,Self}^*$ at time $t = 0$, given physician discharge strategies \underline{t}^* and \bar{t}^* for low- and high-type physicians, respectively. Given this assignment policy, physicians choose the optimal discharge strategies \underline{t}^* and \bar{t}^* .

Proposition 7. *In the Perfect Bayesian Equilibrium in the self-managed system with no physician reporting but with commitment to an assignment policy, if as $\psi > 0$, $\Delta K_P > \bar{K}_N$, and $D = 1$, then there will be less foot-dragging and more ex post efficient assignment than in the nurse-managed system with no physician reporting but triage nurse commitment.*

The scenario for no physician reporting and policy commitment is analyzed similarly as in the corresponding scenario in the nurse-managed system. For any given policy $\pi_{C,Self}$, a low-type physician will discharge his patient at time

$$\underline{t}^* = \underline{\theta} + (1 - \psi) \frac{K_P}{2(\bar{\theta} - \underline{\theta})} \left(\pi_{C,Self} - \frac{1}{2} \right). \quad (1.B.16)$$

Note again the similarity between Equations (1.B.6) and (1.B.16). The only difference is that the second term is multiplied by $1 - \psi$, because with probability ψ , foot-dragging will have no effect on assignment. Self-management – both the observation of true workload and the use of this information in assignment – decreases foot-dragging relative to the nurse-managed system.

It now follows that the physicians will choose a policy function that takes Equation (1.B.16) into consideration in order to maximize their *ex ante* utilities, before they have received their initial patients. They expected to be type $\underline{\theta}$ with probability p and type $\bar{\theta}$ with probability $1 - p$, and they maximize

$$\begin{aligned} \mathbb{E}[u^P(\pi_{C,Self})] &= -p(\underline{t}^* - \underline{\theta})^2 - (1 - p)^2 \Delta K_P - \\ &\quad 2p(1 - p)(1 - \psi) \Delta K_P \left[\frac{1}{2} \left(\frac{\underline{t}^* - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + (1 - \pi_{C,Self}) \left(\frac{\bar{\theta} - \underline{t}^*}{\bar{\theta} - \underline{\theta}} \right) \right], \end{aligned} \quad (1.B.17)$$

where I conveniently transform the expected utility by subtracting \underline{K}_P and using $\Delta K_P \equiv \bar{K}_P - \underline{K}_P$. Again, note the similarity with Equation (1.B.7), with two differences. First, with probability ψ , the policy function is irrelevant for assignment, so $1 - \psi$ appears in the third term.⁵⁷ Second, instead of \bar{K}_N , the analogous parameter ΔK_P from the physician's utility function is used because physicians

⁵⁶For simplicity and for consistency with the nurse-managed system, I assume that patients are immediately assigned under this policy. See also footnote (54).

⁵⁷The third term also represents the efficiency loss with respect to misassignment under any policy commitment. It is larger with small ψ , p close to $1/2$, and large ΔK_P . On the other hand, the efficiency loss with no policy commitment is larger with large k , large w , or p close to 0 or 1. Depending on these parameters, physicians may opt for no policy function even if they can commit to one.

are the ones making patient assignment.

Ex ante, physicians would like to avoid assigning the new patient to a busier physician because it is more costly, by ΔK_P , for that physician to deal with that patient. However, *ex post*, once physicians know their types, low-type physicians have the moral hazard to avoid the new patient. Commitment to a policy function allows physicians *ex ante* to balance their desire for proper assignment with the knowledge that proper assignment will cause costly moral hazard.

Maximizing (1.B.17) with respect to $\pi_{C,Self}$, after substituting (1.B.16) for \underline{t}^* , yields the optimal policy function

$$\pi_{C,Self}^* = \frac{1}{2} + \frac{4p(1-p)\Delta K_P(\bar{\theta} - \underline{\theta})^2}{(1-\psi)\underline{K}_P(p\underline{K}_P + 4p(1-p)\Delta K_P)}. \quad (1.B.18)$$

The differences between (1.B.8) and (1.B.18) are twofold. First, the denominator is multiplied by $1 - \psi$, which reflects the fact that foot-dragging is lessened by the possible observation of true workload, and which improves the efficiency of assignment even when true workload is not observed. Second, ΔK_P replaces \bar{K}_N . Thus, as long as $\psi > 0$ and $\Delta K_P > \bar{K}_N$, then $\pi_{C|D=1}^* < \pi_{C,Self}^*$.

Although I cannot definitively say that one is bigger than the other, it is likely that $\Delta K_P > \bar{K}_N$. To see this, notice that the triage nurse's utility function in Equation (1.B.2) can include the outcomes of *both* physicians. \bar{K}_N represents the amount that she values assignment of the third patient compared to the amount that she values the average of outcomes for the first two patients. On the other hand, in the physician utility function in Equation (1.B.1), the cost of receiving another patient is scaled relative to the outcome of a single patient. So if the triage nurse had similar preferences as both physicians, we should have $\Delta K_P \approx 2\bar{K}_N$. For both of these reasons, $\pi_{C,Self}^*$ again should be greater than π_C^* , meaning that the self-managed system improves the efficiency of assignment relative to the nurse-managed system.

1.B.3.3 Physician Reporting, Policy Commitment

Finally, in the case of physician reporting and policy commitment, it suffices to consider how the observation of true workloads between peers and its use in the policy function modifies incentive compatibility constraints for a low-type physician.

The standard incentive compatibility constraint, assuming “full” lying, is

$$-(t(\bar{\theta}) - \underline{\theta})^2 - \frac{1}{2}(1-\psi)\underline{K}_P \leq -\pi(1-\psi)\underline{K}_P. \quad (1.B.19)$$

Even if he mimics a high-type physician, he will still be observed as a low-type physician with probability ψ , in which case mimicry was useless. This relaxes the incentive compatibility constraint that was originally (1.B.10) and allows a higher π to support truth-telling.

However, as in the nurse-managed system, the intertemporal incentive compatibility constraint will be binding. Recall that this constraint, shown in Equation (1.B.12) for the nurse-managed system, derives from the concern that low-type physicians can lie about their type and then discharge

their patient earlier than $t(\bar{\theta}) = \bar{\theta}$. The constraint in the self-managed system is

$$-(\hat{t} - \underline{\theta})^2 - (1 - \psi) \underline{K}_P \left[\frac{1}{2} \left(\frac{\hat{t} - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + \frac{\bar{\theta} - \hat{t}}{\bar{\theta} - \underline{\theta}} \right] \leq -\pi(1 - \psi) \underline{K}_P, \quad (1.B.20)$$

where \hat{t} is the time that a low-type physician reveals that he was lying when he initially reported that he was high-type. Again, with probability ψ , this lie will not pay off, which relaxes the incentive compatibility constraint to allow a higher π . Thus, assignment will be more efficient in the self-managed system compared to the nurse-managed system in this case as well.

Proposition 8. *In the Perfect Bayesian Equilibrium in the self-managed system with physician reporting and commitment to an assignment policy, assignment will be more ex ante (and ex post) than in the nurse-managed system with physician reporting and triage nurse commitment. There is still no foot-dragging, given truth-telling and discrete types.*

1.B.4 Continuous Types

In this subsection, I relax the baseline two-type model to consider a continuum of types in the nurse-managed system with physician reporting. The purpose of this extended model is to communicate the intuition that, in contrast to the two-type analysis in Section 1.B.2.4 with physician reporting, there will still be foot-dragging and inefficient assignment, as long as types are sufficiently rich. While I restrict attention to the nurse-managed system, similar intuition follows for the self-managed system.

The model is identical to the model outlined in Section 1.B.1 except for two differences: First, each of the two physicians can be of type $\theta_j \in [\underline{\theta}, \bar{\theta}]$, drawn from some distribution which I do not specify.⁵⁸ Second, because types lie in a continuum, I allow for a more flexible triage nurse policy function that takes the form of $\pi(\theta_1, \theta_2, t)$, which is the flow probability of the physician 1 receiving a patient who arrives at time t when the types of physicians 1 and 2 are θ_1 and θ_2 , respectively. Of course, this policy function cannot be reduced to a single parameter, because both θ_1 and θ_2 are continuous. In addition, I allow for the fact that the optimal policy may have a non-constant flow-rate for a given θ_1 and θ_2 . The sufficient statistic from the standpoint of physician and triage nurse utility is

$$P(\theta_1, \theta_2) \equiv \frac{1}{\bar{\theta} - \underline{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \pi(\theta_1, \theta_2, t) dt,$$

which is the cumulative probability over time that physician 1 will receive the new patient, given that patient arrival is uniformly distributed and that assignment is immediate upon arrival. The reason I allow for the flexible time-dependent flow is to specifically consider “intertemporal feasibility” constraints in which $\pi(\theta_1, \theta_2, t) \leq 1$, for all θ_1, θ_2, t , which I have noted in Section 1.B.2.4 as feature

⁵⁸The distribution is not important for this analysis, which shows positive foot-dragging and inefficient assignment, but it would be necessary to consider for an analysis that computes the optimal assignment function in closed form.

different from the standard screening problem.

I will analyze this model as follows: First, I will formalize the intertemporal feasibility constraints as incentive compatibility constraints for truth-telling. That is, physicians should have no incentive to discharge their patient at time $t < t(\hat{\theta}_j)$, where $\hat{\theta}_j$ is their reported type, which may be different than their true type θ_j . Second, as in the standard mechanism design problem, I will show that I can restrict attention to local incentive compatibility constraints in which physicians have no incentive to report the type continuously adjacent to theirs, conditional on the previous requirement that they follow the discharge time required by that report. Third, I will show by perturbation arguments that the optimal triage nurse policy function is continuous and strictly increasing. This implies that there will be positive foot-dragging and *ex post* inefficient assignment in the sense that $P(\theta_j, \theta_{-j}) < 1$ for any $\theta_j < \theta_{-j}$. That is, in the remainder of this subsection, I will show the following:

Proposition 9. *In the Perfect Bayesian Equilibrium in the nurse-managed system with physician reporting and triage nurse commitment, for a continuum of physician types distributed along $[\underline{\theta}, \bar{\theta}]$, there will be positive foot-dragging such that $t(\theta) > \theta$ for all $\theta < \bar{\theta}$ and ex post inefficient assignment such that $P(\theta_j, \theta_{-j}) < 1$ for any $\theta_j < \theta_{-j}$.*

1.B.4.1 Intertemporal Feasibility

As in any truth-telling equilibrium, physicians must not have the incentive to misreport their types. This setting in particular requires me to address the possibility of a physician reporting $\hat{\theta}_j$, in order to get some flow probability $\pi(\hat{\theta}_j, \theta_{-j}, t)$ of assignment for the new patient, but discharging the current patient at some time $t_j < t(\hat{\theta}_j)$. If physicians can receive a lower $P(\hat{\theta}_j, \theta_{-j})$ by reporting $\hat{\theta}_j > \theta_j$ but not have to keep their patient as long as would be required by $t(\hat{\theta}_j)$, then they could be strictly better off. The reason for this departure from the standard screening model is that the cumulative probability $P(\theta_1, \theta_2)$ is not given by the triage nurse in a lump sum but rather over time. Thus, there is an “intertemporal feasibility” constraint in that the triage nurse may not be able to take back (in terms of lower probability of assignment) what she has already given in the past. More precisely, this constraint derives from the fact that even punishment policy functions are limited by $\pi(\theta_1, \theta_2, t) \leq 1$, for all θ_1, θ_2 , and t .

In order to account for this, I simply consider that, for the standard mechanism design problem to work here, I need physicians to have no incentive to report $\hat{\theta}_j$ and then discharge their patient at time $t_j < t(\hat{\theta}_j)$ as opposed to reporting $\tilde{\theta}_j$ and discharging their patient at the same time $t_j = t(\tilde{\theta}_j)$. That is, if the physician is planning to discharge his patient at some time t_j , he may as well report the type that corresponds to that time. Note that this incentive compatibility does not yet require that the physician prefers to report $\tilde{\theta}_j = \theta_j$, which I consider later. In order to sustain this aspect of truth-telling, I assume that if the triage nurse catches a physician lying *and* deviating by $t_j < t(\hat{\theta}_j)$, then she will punish him by assigning him the new patient with probability 1 – but no more by the feasibility constraint – if the new patient arrives at $t \in [t_j, \bar{\theta}]$.

The incentive compatibility constraint implied by intertemporal feasibility then can be stated as

$$\begin{aligned} P(\tilde{\theta}_j, \theta_{-j}) &\equiv \frac{1}{\bar{\theta} - \underline{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \pi(\tilde{\theta}_j, \theta_{-j}, t) dt \\ &\leq \frac{1}{\bar{\theta} - \underline{\theta}} \left[\int_{\underline{\theta}}^{t(\tilde{\theta}_j)} \pi(\hat{\theta}_j, \theta_{-j}, t) dt + \bar{\theta} - t(\tilde{\theta}_j) \right], \forall \hat{\theta}_j < \tilde{\theta}_j, \theta_{-j} \end{aligned}$$

where I ignore the potential cost of discharging a patient at time $t(\tilde{\theta}_j)$, since this is held constant. This incentive compatibility constraint can equivalently be stated as

$$\int_{t(\theta_j)}^{\underline{\theta}} 1 - \pi(\theta_j, \theta_{-j}, t) dt \geq \int_{\underline{\theta}}^{t(\theta_j)} \pi(\theta_j, \theta_{-j}, t) - \pi(\hat{\theta}_j, \theta_{-j}, t) dt, \forall \hat{\theta}_j < \theta_j, \theta_{-j}. \quad (1.B.21)$$

This formalizes the intuition that the triage nurse requires scope for punishment in the policy function in order to prevent physicians from misreporting their types.

1.B.4.2 Local Incentive Compatibility

The remainder of the analysis proceeds similarly to the standard screening mechanism design. The triage nurse faces a problem in which she offers a menu of choices $\{(t(\theta_1, \theta_2), P(\theta_1, \theta_2))\}$ to physicians. Physicians simultaneously report θ_1 and θ_2 , and by the Revelation Principle, they report truthfully.⁵⁹ The menu that the triage nurse chooses to offer maximizes her expected utility subject to physician truth-telling.

I first show that the single-crossing condition holds. Physician utility remains the same as in the baseline model, but I account for continuous types by allowing the cost of a new patient to a physician of type θ , $K_P(\theta)$, to be a continuous and differentiable function such that $K'_P(\theta) \geq 0$. It is easy to show that

$$\frac{\partial}{\partial \theta_j} \left[- \frac{\partial u_j^P / \partial P(\theta_j, \theta_{-j})}{\partial u_j^P / \partial t_j} \right] > 0$$

as long as $t > \theta$ and $K'_P(\theta) \geq 0$. The former condition that $t > \theta$, equivalent to stating that there is positive foot-dragging, will be shown later to hold in equilibrium. I will assume the latter more strictly by $K'_P(\theta) > 0$.

Given the single-crossing condition, it is then sufficient to summarize the set of incentive compatibility constraints,

$$-(t(\theta_j, \theta_{-j}) - \theta_j)^2 - K_P(\theta_j) P(\theta_j, \theta_{-j}) \geq -(t(\hat{\theta}_j, \theta_{-j}) - \theta_j)^2 - K_P(\theta_j) P(\hat{\theta}_j, \theta_{-j}),$$

⁵⁹In the following analysis, I assume that physicians know each other's types, but the intuition follows if they only act according to expected values of their peer's type.

for all $\theta_j, \theta_{-j}, \hat{\theta}_j$, as a monotonicity condition and local incentive compatibility constraints. The monotonicity condition is

$$\frac{\partial P(\theta_j, \theta_{-j})}{\partial \theta_j} \leq 0, \quad (1.B.22)$$

and the local incentive compatibility constraints are simply stated by setting the derivative of the physician utility function equal to 0.⁶⁰

$$-2(t(\theta_j, \theta_{-j}) - \theta_j) \frac{\partial t(\theta_j, \theta_{-j})}{\partial \theta_j} = K'_P(\theta_j) \frac{\partial P(\theta_j, \theta_{-j})}{\partial \theta_j}. \quad (1.B.23)$$

1.B.4.3 Optimization Problem

I will now analyze the triage nurse's optimization problem subject to (1.B.21) and (1.B.23). Rather than solve her problem in closed form, I can obtain my results – that there will be foot-dragging and *ex post* inefficient assignment in equilibrium – by simple perturbation arguments.

I first maintain the assumption that the triage nurse's utility function takes the form as stated in (1.B.2), and I similarly assume that $K_N(\theta)$ is continuous and twice differentiable and that furthermore $K'_N(\theta) > 0$. It then follows that she will offer menu options $t(\theta_j, \theta_{-j})$ and $P(\theta_j, \theta_{-j})$ that are continuous and differentiable in θ_j for all θ_j and θ_{-j} . To see this, suppose that her optimal assignment policy function $P(\theta_j, \theta_{-j})$ is discontinuous at $\theta_j = \theta_1$ and some $\theta_{-j} = \theta_2$, such that $\lim_{\epsilon \rightarrow 0} P(\theta_1 - \epsilon, \theta_2) - P(\theta_1 + \epsilon, \theta_2) = \Delta > 0$. If this is the case, then no physician whose type $\theta_j \in (\theta_1 - \epsilon, \theta_1)$, for some $\epsilon > 0$, will truthfully reveal his type. The triage nurse cannot maintain truth-telling over some interval of types of strictly positive measure and therefore cannot implement her policy function, which is a contradiction.

Next, note that it is never optimal to have $t(\theta) < \theta$, as it reduces the discharge time away from what is socially optimal and only reduces the utility of the physician when truth-telling, which can only make the truth-telling constraint more costly. Given this, I can show that the optimal assignment policy $P(\theta_j, \theta_{-j})$ is strictly decreasing in θ_j . To see this, suppose that the optimal policy is such that $P(\theta_j + \epsilon, \theta_{-j}) \geq P(\theta_j, \theta_{-j})$ for some small $\epsilon > 0$ and for some θ_j, θ_{-j} . However, the triage nurse can strictly increase her utility and maintain truth-telling by increasing $P(\theta_j, \theta_{-j})$ by some small $\epsilon > 0$ while keeping her set of $\{t(\theta_j, \theta_{-j})\}$ unchanged if $\partial t(\theta_j, \theta_{-j}) / \partial \theta_j > 0$, or by increasing $P(\theta_j, \theta_{-j})$ by some small $\epsilon > 0$ and concurrently decreasing $t(\theta_j, \theta_{-j})$ by some small $\delta > 0$ if $\partial t(\theta_j, \theta_{-j}) / \partial \theta_j \leq 0$. Note that this satisfies the monotonicity condition in (1.B.22).

Given that $P(\theta_j, \theta_{-j})$ is strictly decreasing in θ_j (by contradiction), $t(\theta) \geq \theta$ (by contradiction), and $K'_P(\theta_j) > 0$ (by assumption), then (1.B.23) implies that $\partial t(\theta_j, \theta_{-j}) / \partial \theta_j > 0$ and $t(\theta_j) > \theta_j$, except for type $\theta_j = \bar{\theta}$, which need not be bound by an incentive compatibility constraint. That is, discharge times increase with the physician's type, regardless of his peer's type, and there is positive foot-dragging in equilibrium for all types $\theta_j < \bar{\theta}$.

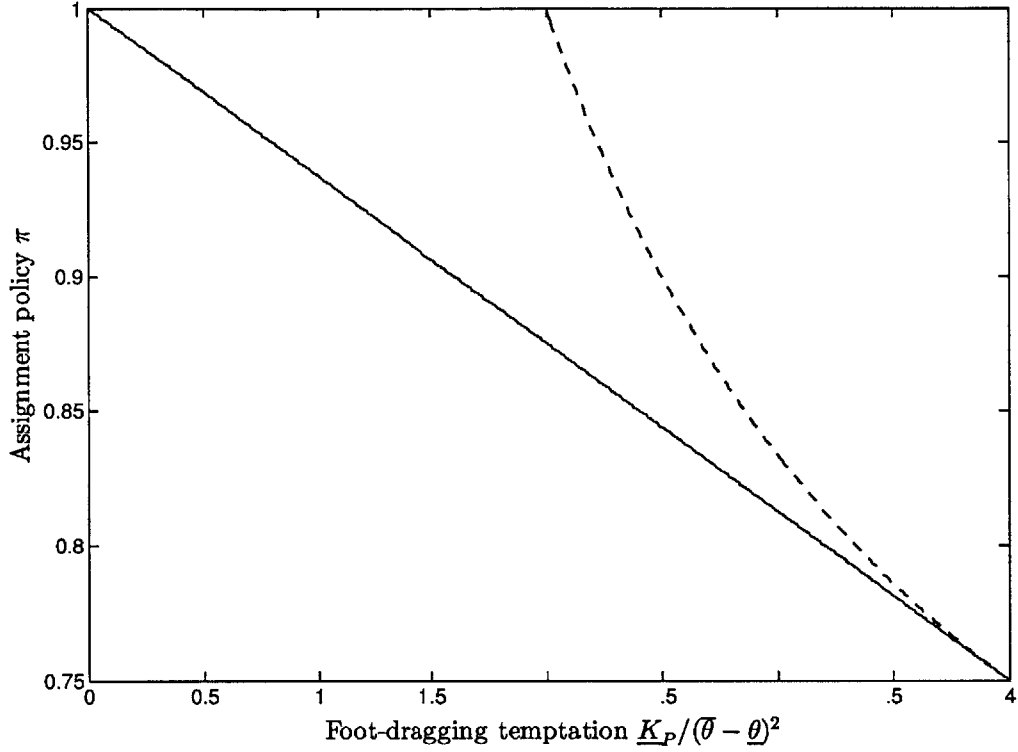
⁶⁰ An additional condition for the local incentive compatibility constraint to be valid is that the menu options $t(\theta_j, \theta_{-j})$ and $P(\theta_j, \theta_{-j})$ are continuous and differentiable in θ_j for all θ_j and θ_{-j} . This will also be shown below.

Finally, *ex post* inefficient assignment derives from two independent facts shown above, either of which is sufficient. First, the continuity of $P(\theta_j, \theta_{-j})$ guarantees *ex post* inefficient assignment because it is impossible to have

$$P(\theta_j, \theta_{-j}) = \begin{cases} 1, & \theta_j < \theta_{-j} \\ 0, & \theta_j > \theta_{-j} \end{cases}, \quad (1.B.24)$$

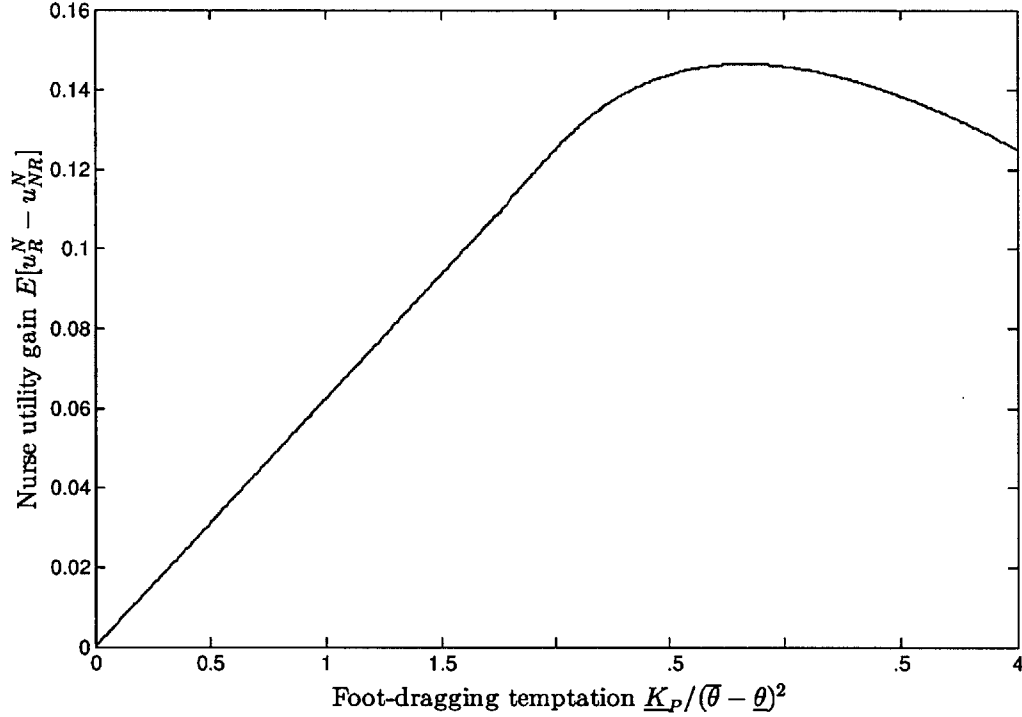
for all θ_j, θ_{-j} , without a discontinuity at $\theta_j = \theta_{-j}$. Second, inefficient assignment can also be proven by using intertemporal feasibility alone, as stated in (1.B.21). This constraint implies that it is impossible to have both $P(\theta_j, \theta_{-j}) > P(\hat{\theta}_j, \theta_{-j})$ and $P(\theta_j, \theta_{-j}) = 1$, for any $\hat{\theta}_j, \theta_j$, and θ_{-j} , and therefore also rules out (1.B.24).

Figure 1.B.1: Assignment Policy π Depending on Physician Reporting



Note: This figure shows the assignment policy π , depending on whether physicians can report their types to the triage nurse. Under both cases, I assume that the triage nurse can commit to an assignment policy function. Under no physician reporting, I also consider the triage nurse utility function in which she only cares about patient assignment. On the horizontal axis is a summary statistic for the foot-dragging temptation, $\underline{K}_P / (\bar{\theta} - \underline{\theta})^2 \in [0, 4]$. Note that when $\underline{K}_P / (\bar{\theta} - \underline{\theta})^2 = 4$, a low-type physician should fully foot-drag at $\underline{t}^* = \bar{\theta}$. In dashed blue, I plot the assignment policy when physicians cannot report, $\pi_{C|D=0}^*$ given in (1.B.9). In solid red, I plot the assignment policy when physicians can report, $\pi_R^* = \pi_{IT}^*$, which equals the assignment policy implied by the intertemporal constraint given in Equation (1.B.13). Note that $\pi_{C|D=0}^* = \pi_{IC}^*$, the assignment policy implied by the standard incentive compatibility constraint with “full” lying, given in Equation (1.B.11); however, the former is a function of censuses, while the latter is a function of reported types.

Figure 1.B.2: Efficiency Gain in *ex ante* Assignment with Physician Reporting



Note: This figure shows the gain in *ex ante* assignment efficiency, or the gain in triage nurse utility when she only cares about assignment, that occurs with physician reporting. I compare expected utility for the triage nurse, $\mathbb{E}[u^N]$, under no physician reporting and under physician reporting. In both cases, her expected utility is simply $\Pr\{\theta_{j(3)} = \bar{\theta}\}$, normalizing $\bar{K}^N = 1$. Under both cases, I assume that the triage nurse can commit to an assignment policy function. On the horizontal axis is a summary statistic for the foot-dragging temptation, $\underline{K}_P / (\bar{\theta} - \underline{\theta})^2 \in [0, 4]$. Note that when $\underline{K}_P / (\bar{\theta} - \underline{\theta})^2 = 4$, a low-type physician should fully foot-drag at $\underline{t}^* = \bar{\theta}$. On the vertical axis, I plot the difference in *ex ante* assignment efficiency between physician reporting and no reporting.

Chapter 2

The Dynamics of Housestaff Physician Practice Styles

2.1 Introduction

An extensive literature has documented substantial and robust variation in medical care between regions and within regions.¹ Although the term “physician practice styles” has been broadly used to describe this variation, the exact mechanisms behind such regional variation have been at best loosely defined. For example, practice styles have been variously attributed to physician beliefs about appropriate care (Phelps, 2000; Phelps and Mooney, 1993; Coleman et al., 1966), local resources or expertise (Chandra and Staiger, 2007), or local cultural norms (Burke et al., 2010).

Despite these broad hypotheses, surprisingly little has been shown about the microfoundations of practice styles, holding the institutional environment constant. Evidence at the institutional level does not account for physician selection, and evidence across physicians may conflate physician beliefs with unobserved institutional resources. Using relatively coarse measures, researchers have found physicians with extremely stable practice styles, even immediately after a move (Molitor, 2011) or starting their practice (Epstein and Nicholson, 2009).²

The question then remains: What is the most important mechanism behind practice-style variation for physicians sharing the same institutional resources? Physicians are not born knowing how to practice medicine. Yet, measured practice styles fail to converge for physicians in the same area. This could be due to many reasons, some reflecting study limitations and others behavioral mechanisms: First, studies do not account for unobservable institutional constraints. Physicians may also choose to practice and stay in institutions where their practice styles are consistent with

¹Papers in this literature include Wennberg et al. (2002); Wennberg and Gittelsohn (1973); Wennberg et al. (2004); Fisher et al. (2003b,a). More recently, Epstein and Nicholson (2009); Grytten and Sorensen (2003) have shown that variation within regions is potentially greater than that across regions.

²Molitor (2011) finds that practice styles change with a move but does not find any continued dynamics after the move.

local norms or resources. Second, coarse or infrequent measures cannot capture dynamics that occur within relatively short periods of time. Third, practice styles may reflect preferences, abilities, or other unchanging characteristics independent of medical knowledge. Finally, practice styles could be formed mostly in residency training, prior to being observed in essentially any study, to my knowledge, of practice-style variation.

In this paper, I aim to study the dynamics of practice styles of physicians in internal medicine residency training, or “housestaff” physicians, at a single, large academic hospital. I observe housestaff physicians throughout the entire course of their training. They start training with extremely little clinical experience, and they finish training prepared and eligible to practice medicine without supervision. By construction, housestaff physicians are exposed to *ex ante* similar clinical experiences, but of course, due to scheduling, they experience *ex post* different histories of patients, coworkers, supervisors, and clinical rotations at the same relative point in their training. This provides an opportunity to examine the effect of *learning* from training experiences – some observed and some unobserved or difficult to summarize – on practice styles.

Inpatient care by housestaff physicians is explicitly delivered in teams. Although this increasingly reflects the reality of clinical care anywhere, it is usually overlooked in the existing literature on physician practice styles, in part because members of the clinical team are usually unobserved. In contrast, I not only observe all physicians in the clinical team, exogenously assigned to each other, I also observe the same housestaff physicians discontinuously changing job roles as they move forward in training.³ First-year housestaff, or “interns,” and second- and third-year housestaff, or “residents,” share patients and have the same formal job descriptions.⁴ Both interns and residents recommend and order tests and treatments. However, in an organizational culture that respects hierarchy, the opinions of residents are followed much more than those of interns. This constitutes a unique setting to examine the effect of *authority*, or the degree to which physician beliefs and preferences influence what is actually done.

Finally, I observe a rich set of housestaff physician characteristics, with which I can examine physician groups that might be expected *a priori* to have different practice styles due to *intrinsic heterogeneity*. First, I use the fact that some housestaff are predetermined to stay in internal medicine training for only one year, prior to starting another residency in a field such radiology or dermatology. Second, other housestaff on the same clinical services rotate from another hospital, both as interns and as residents, or from another training program such as emergency medicine or obstetrics-gynecology. Third, due to the detailed selection process by the residency program, I observe not only housestaff physician demographics but also measures of pre-training ability, such as United States Medical Licensing Examination (USMLE) test scores, and desirability, such as

³I also observe the supervising “attending” physician, whom I will describe later in Section 2.3.

⁴All housestaff physicians can also be generically called “residents,” but I will avoid this use in this paper in order to avoid confusion. More specific names for second- and third-year housestaff include “junior residents” and “senior residents,” respectively, or “supervising residents,” but I will use “residents” for brevity given that there will be no ambiguity in this use.

positions on the residency program's rank list.

I use radiology and laboratory testing as a measure of practice styles. Testing costs have been a focus both in practice-style variation and in cost control, particularly as the relative costs of testing have increased and now comprise a significant portion of overall spending (Bates et al., 1999; Schroeder et al., 1974; Iwashyna et al., 2011). Importantly for this study, tests are frequently ordered and observed in the data, allowing for precise estimation of practice styles within short timeframes. Throughout this study, I define practice styles in terms of the effect of a physician on what is actually done in clinical care, which makes clear that practice styles can be a combination of intrinsic heterogeneity, learning, and authority.

My results are the following: First, I find that housestaff physicians are responsible for significant variation in test spending at any time in their training. Interns who are one standard deviation above the mean incur 15-30% higher spending costs. Residents who are one standard deviation above the mean incur 55-80% higher spending costs. The fact that variation attributable to housestaff physicians actually increases from interns to residents, despite similar average experiences, is inconsistent with learning as the sole mechanism explaining practice styles. Rather, I find that the variation in practice styles increases discontinuously as interns become residents, suggesting that authority plays a significant role in practice-style variation.

Second, I study the serial correlation of practice styles across short periods of time in training, and I find that, far from being static, practice styles change throughout training and until its very end. While practice styles remain unstable during the first year of training, they rapidly become very stable after interns become residents, specifically after housestaff are given more authority. Learning thus appears to depend on both experience and authority. However, some evidence suggests that practice styles become less stable when patients are seen less frequently, for example in the third year of training, or when equally spaced clinical rotations are less contiguous in between. In other words, there seems to be both learning and forgetting.

With respect to learning, I find very little evidence that easily summarizable experiences during training influence housestaff practice styles to any significant degree. In particular, rotations as varied as outpatient care, research, subspecialty care, and intensive care unit (ICU) time within the same hospital do not appear to have a significant effect on practice styles. The same is true for time working with supervising physicians who are measured to have high- or low-spending practice styles. Learning thus seems to be highly individualized, allowing for variation despite similar experiences. However, I do find that time spent at an affiliated community hospital reduces test spending by more than half, suggesting that experiences across institutions, consistent with institutional norms, are most important for explaining differences in learned practice styles.

Finally, I explore the role of intrinsic heterogeneity in determining practice styles, to the extent that such heterogeneity is correlated with characteristics predating residency training or choices made prior to starting residency. For a wide variety of housestaff categories, I find very little relation with practice styles. I do find that housestaff who are male or those who have higher

USMLE scores spend slightly less. Interestingly, these effects grow with tenure, suggesting that time-invariant heterogeneity does not automatically lead to practice-style differences but must be allowed to manifest, through authority. In contrast, residents visiting from another hospital with an emphasis on cost control spend 17% less, while interns from this other hospital incur no different spending, compared to respective residents and interns from the main hospital. While this could reflect selection into the other hospital, its magnitude with respect to other housestaff characteristics provides further support for the importance of learning institutional norms, magnified by authority.

These findings support a dynamic picture of physician practice styles that involves both learning and authority. As physicians gain experience, their practice styles become more stable but settle at divergent places. An increase in informal authority, as housestaff progress from interns to residents, is responsible for at least a doubling in the standard deviation of practice styles and an acceleration in the stability of practice styles. Given similar experiences, the fact that practice styles can be so different suggests that physician learning is highly individualized. As described by Gawande (2012), “[physicians] learn what they want to, when [they] want to.” Nonetheless, experiences across institutions have a substantial and lasting impact, highlighting the importance of institutional norms and suggesting scope for cultural change.

The remainder of the paper is organized as follows. Section 2.2 lays out a conceptual framework with which to consider intrinsic heterogeneity, learning, and authority as influences on observed practice styles. Section 2.3 describes the institutional setting, and Section 2.4 describes the data. Section 2.5 presents the main results on the practice-style distributions and dynamics over time; Section 2.6 discusses whether major categories of observable learning experiences affect practice styles; and Section 2.7 tests for the dependence of practice styles on intrinsic heterogeneity associated with housestaff characteristics. Section 2.8 concludes and discusses policy implications.

2.2 Conceptual Framework

In this section, I describe a conceptual framework of practice styles that includes three main components – intrinsic heterogeneity, learning, and authority. In this paper, practice styles are defined as the effect of a physician on actual patient care. This is implicitly the approach taken in the existing literature on practice variation, but the theoretical justification for explaining such variation has often been vague or incomplete. For example, much emphasis has been placed on learning as a mechanism that would generate practice-style variation, for example through “schools of thought” (Phelps and Mooney, 1993), but in practice, even physicians exposed to the same training do not prefer to do the same thing. It is also difficult to imagine a physician with any belief always being allowed to do what he wanted regardless of setting; modern institutions of health care delivery simply do not work that way.

These components of practice styles are not mutually exclusive and in fact likely interact with each other. For example, at a given institution, physicians with certain preferences or beliefs are

likely to be given more authority than others. The effect of learning may also differ depending on preferences. Nonetheless, they are useful as separate concepts for interpreting and categorizing empirical results.

2.2.1 Intrinsic Heterogeneity

Any realistic framework of physician practice-style variation must consider intrinsic heterogeneity. Even with perfect knowledge, physicians differ in their risk aversion, skills, patience, personality, and relative valuation of risks and benefits. Of course, knowledge in medicine is not perfect, even in textbooks. Medical decisions rarely fit into clean buckets of evidence from randomized controlled trials without conflicting with other information, from competing clinical trials, physiologic deduction, clinical intuition, or patient preferences. Medicine is predominantly an individual endeavor between a physician and a patient, rather than a population-level policy, and any number of factors must be weighed in limited time using rules of thumb. Such uncertainty and limitations on rationality yield even further scope for intrinsic heterogeneity among physicians to weigh on practice styles.

Although heterogeneity is usually treated as an assumption independent of the decision context, recent work in psychology and behavioral economics has also underlined the possibility that decision-makers are often unsure of their specific preferences but could use context and generic heterogeneity to settle on a choice (Kamenica, 2008). Specifically, when faced with a wide variety of choices, a given physician may always pick the “conservative” choice, the “aggressive” choice, or the “standard” choice.

I observe a number of housestaff characteristics and choices that are predetermined prior to residency training and likely correlated with heterogeneous preferences or skills. Characteristics include demographics, performance on standardized tests, and relative desirability in the residency selection process. Choices include specific residency tracks that could involve identical training in the first year but are scheduled to diverge after that.

2.2.2 Learning

From a conceptual perspective, physicians in training may learn from actual clinical knowledge gained idiosyncratically by caring for patients, from supervising physicians through shared experience (Van den Steen, 2010) even if no knowledge is gained (Banerjee, 1992), and from institutional norms in the form of “word-of-mouth” communication and rules of thumb (Ellison and Fudenberg, 1993). The key here is that learning need not be the literal acquisition of true information but can occur merely by working with someone else with a given clinical belief or by being in an environment where a given belief is accepted as the right way to practice medicine.

I identify specific learning experiences by the history of clinical rotations within the same hospital, which involve different types of patients; supervising physicians; and rotations to an affiliated community hospital. The latter is likely to represent a more disparate learning experience because

it includes differences in institutional norms in addition to the other channels of learning. It is important to note that, with initial heterogeneity, all types of learning should result in convergence in practice styles if experiences are sufficiently many, because residency exposes housestaff to the same training experiences *on average*.⁵

2.2.3 Authority

Finally, the profession of medicine is fundamentally one of expertise. If patients or health care administrators could easily determine whether medicine is practiced correctly by a cookbook, then there should be no variation in care under the same patient or administrator preferences. But as physicians gain expertise, they also gain authority to implement decisions potentially away from the norm, and practice styles may diverge.

Relatedly, medicine is not practiced in isolation. Physicians routinely consult with each other, coordinate with ancillary staff, and operate within the bounds of a greater delivery system. Physicians can be constrained literally by capacity limits on utilization or more subtly by defaults or requirements for consultation. More broadly, physicians with greater knowledge, stature, or communication skills are less likely to be questioned by nurses or reported to their superiors as practicing out of guidelines. Finally, as inpatient care is increasingly delivered by teams of physicians, even in non-academic settings, physicians with greater authority are better able to persuade their colleagues to accept a given style of care; in contrast, physicians with lesser authority sometimes do not even voice their opinions.

Authority in decision-making has been theoretically explored in principal-agent settings with incomplete principal information and biased agent incentives with respect to choices (Crawford and Sobel, 1982; Aghion and Tirole, 1997). More generally, however, partial authority need not require systematic bias or different preferences, but merely incomplete information about the optimal decision by both parties and a desire to be listened to (or multidimensional payoffs) by the agent (Che and Kartik, 2009; Rantakari, 2011; Scharfstein and Stein, 1990; Ottaviani and Sorensen, 2001).

In practice, authority is likely to depend on preferences, ability, and beliefs. However, I exploit a unique feature of my setting in that I observe the same housestaff physicians changing roles from intern to resident, with a discontinuous increase in authority. I will further discuss this institutional feature in the next section, but I note here that I exploit the plausible assumption that preferences and ability are continuous across this change in role, and that there are no mechanical differences in the ability to order tests between the two roles.

⁵Learning that involves only a subset of experiences, such as habit formation or forgetting, may involve less-than-full convergence, but again if practice styles are determined only by learning, then variation should still weakly decrease over time.

2.3 Institutional Setting

I study the development of practice styles among internal medicine physicians in training at a large academic medical center. Academic medical centers are responsible for training new physicians through residency programs that accept graduates from medical school. In residency, a tremendous amount of clinical experience is acquired. The average graduating medical student may have taken care of less than a dozen patients in internal medicine, while by the end of residency, he will have personally admitted hundreds and participated in the care of well over a thousand patients. Physicians are often judged by where they “trained,” i.e., at which hospital they completed their residency, rather than what medical school they attended. It is therefore no surprise that academic medical centers are largely viewed as the source of divergent practice styles (Phelps and Mooney, 1993; Wennberg et al., 2004), although simply considering the place of residency training does not explain much of the variation in practice styles (Epstein and Nicholson, 2009). To my knowledge, the development of practice styles within residency has not yet been studied.

Regulated by the American Board of Internal Medicine (ABIM) and the Accreditation Council for Graduate Medical Education (ACGME), internal medicine residencies are roughly the same in structure across the United States. The number of clinical weeks, hours per week, patients per year, and types of rotations are all stipulated by regulation. Physicians in training, also called “housestaff” physicians, generally train for three years in internal medicine. During residency training, they are supervised by physicians who have finished training, called “attending physicians.” That is, each patient is shared by housestaff and attending physicians. While housestaff may document their findings and order tests and treatment, attending physicians are legally responsible for the actions of housestaff under their supervision and are required to attest agreement with assessments and treatment plans laid out by housestaff.

Housestaff in their first year of training are called “interns;” those in their second or third year of training are called “residents” (“junior residents” or “senior residents,” respectively). Within residency, interns and residents work in teams with attending physicians. Each resident is usually assigned to two interns. Depending on the hospital, there are some cultural nuances that set expectations for different roles between interns and residents. In general, the intern is considered to be the “responding clinician,” in that he is usually the person that answers pages from nurses, provides detailed documentation, and enters most orders. Residents are meant to supervise the interns, answering frequent and routine clinical questions. The differences between intern and resident roles are not formalized, however. Legally, all housestaff are treated equally as physicians who have not yet completed their training. Residents may answer pages from nurses, write notes, and directly enter orders. Interns may also unilaterally make decisions without having to go through residents (or attendings). However, the culture of academic medicine is usually collaborative. Other than the internal medicine housestaff and the attending physician, it is not unusual to have several other physicians involved in the care of a single patient, including subspecialty consultations and

overnight physicians. Important decisions are usually discussed before implemented.

In this academic medical center, housestaff from different programs or different “tracks” within a program work together on the same clinical services. For example, a sizeable number of interns only plan to spend one year in the internal medicine residency (“preliminary” interns, as opposed to the standard “categorical” interns), subsequently proceeding to other residency programs, such as anesthesiology, radiology, or dermatology.⁶ These residency plans are committed to prior to starting the internal medicine residency. Internal medicine interns and residents from another academic hospital and non-internal-medicine interns from the same hospital also rotate through the same internal medicine services. The other hospital is anecdotally known to be less research-oriented, slightly lower-ranked in prestige, and with a greater emphasis on cost control for inpatient care. The other residency programs within the same hospital that rotate through the internal medicine services are obstetrics-gynecology and emergency medicine.

Also, housestaff from the main residency program rotate onto medicine blocks at an affiliated community hospital. At this community hospital, the patient population presents with more common illnesses and are generally less complicated, attending physicians who only practice in the community are less interested in teaching textbook medicine or considering lengthy differential diagnoses, and subspecialty consultations are more sparsely staffed. In addition, resources for testing are limited relative to the main academic hospital. Times for blood draws are relatively limited, and nurses might not know how to collect other less common samples. Less common laboratory tests might either be unavailable or require sending out to a distant laboratory for processing. It may take longer for radiology tests to be performed or to have their reports appear in the record, and radiologists are more difficult to contact for consultation.

Conditional on tracks, housestaff rotations are exogenously assigned, and conditional on rotations, exposure to patients and coworkers are exogenously assigned relative to housestaff identities.⁷ Housestaff schedules are arranged a year in advance to satisfy hospital programmatic requirements and broader regulations. Housestaff do not state preferences about the composition or order of their rotations, which include intensive care unit (ICU), outpatient, research, subspecialty, and ward blocks. This study focuses on ward rotations, which comprise of general (inpatient) medicine, including general medicine at the community hospital, cardiology, and oncology blocks. Per the residency administration, preferences are not collected about rotations, and assignment is also not made according to housestaff characteristics, although housestaff on certain tracks may be unavailable during certain times because of responsibilities at other institutions.⁸ It is also exceedingly rare

⁶In addition, there are other much smaller special tracks, including primary care, “short tracks” to subspecialty fellowship training, research tracks such as genetics, and medicine-pediatrics or medicine-psychiatry combined programs.

⁷I discuss formal evidence of this in the next section and in the Appendix. Here I describe the institutional setting that is consistent with this statement.

⁸Housestaff are allowed to express preferences about vacation days, although these vacation days are few, about two weeks per year. Senior residents may also express more general preferences about the timing of non-clinical blocks, such as research electives. For interns, schedules are assigned even prior to their arrival from possibly distant

for housestaff to trade blocks, given that programmatic and regulatory requirements must be met for each housestaff physician, and because scheduling is very difficult for the administration to redo. Scheduling does not consider the teams of intern, resident, and attending physicians that will be formed as a result; in fact, attending schedules are done independently, and neither housestaff nor attending scheduling is aware of each other's results in advance. Housestaff blocks are two weeks in length and staggered for interns and residents. As a result, housestaff teams change at least every week, and it is common that care for a single patient is assumed by multiple teams.

Medical spending has been the focus of much of the literature on practice-style variation (Fisher et al., 2003b,a) and is a key policy focus in its own right (Anderson et al., 2005). I study laboratory and radiology test spending in particular. Test spending has received increasing attention as the relative cost of tests has risen and now comprises a significant proportion of overall costs (Bates et al., 1999; Schroeder et al., 1974; Iwashyna et al., 2011). In this academic medical center, test costs comprise 10% of overall costs, which includes costs for physician and nurse salaries and operating costs. In addition to increasing in size, test costs are also possibly the component under the greatest control by physicians and potentially even more by housestaff physicians (Iwashyna et al., 2011). Patients do not need to consent to tests other than HIV testing, and they generally care less about tests, which they view as having no risk or cost, than about treatment. Compared to treatment, evidence-based guidelines are largely silent on best practices for testing.

2.4 Data

This study uses data collected from several sources. First, I observe the identities of each physician on the clinical team – the intern, resident, and attending physician – for each patient on an internal medicine ward service and for each day in the hospital. Over five years, I observe complete data for 48,185 admissions. Corresponding to these admissions are 724 unique interns, 410 unique residents, and 540 unique attendings. Of the housestaff, 516 interns and 347 residents are from the same-hospital internal medicine residency, with the remainder visiting from another residency program within the same hospital or from the other hospital. The vast majority of housestaff physicians are observed in the data as both an intern and a resident.⁹

The mean number of admissions for interns on the ward services of interest is 106; this includes numbers of admissions for visiting interns from the other hospital, which of course are much fewer than same-hospital interns. The corresponding mean number of admissions for all residents, including visiting residents, is 159. Residents of course see patients over two years (the second and third years of training), while internship is only one year long. Residents see fewer patients in their third year, as their clinical skills have presumably solidified and as time is allowed for research or subspecialty electives, in order for some to prepare for careers after residency. Thus the mean

medical school and prior to meeting the residency administration.

⁹This would not occur, for example, for an intern that started during the last year of my data or for a resident who finished during the first year of my data.

number of admissions for second-year residents is 129, while the mean number is 77 for third-year residents. Attending physicians have a much more varied patient load than housestaff, as they can be primarily clinical or research, and as they can see patients in more than one hospital. The median number of patients seen by an attending over five years is 187, while the 90th percentile attending sees 627 patients over five years.

I observe detailed cost information corresponding to every order that a physician might place. I restrict attention to test costs, which I define as any cost incurred by a radiology (e.g., CT, MRI, nuclear medicine, ultrasound) or laboratory test order. I aggregate these for each patient-day, resulting in 220,117 observations, each linked to a unique intern, resident, and attending physician. The distribution of daily test costs is heavily right-skewed. I censor daily test cost observations greater than \$800, which comprise 3% of the data; the resulting distribution is shown in Figure 2.4.1.¹⁰ The mean daily test cost is \$124, while the median is \$49 and the 90th percentile is \$337. These daily costs aggregate to overall admission tests costs with a mean of \$714.

For each patient admitted to a ward service, I observe detailed demographic and clinical information. Demographics include patient age, sex, race, and language. Clinical information derives primarily from billing data, in which I observe International Classification of Diseases, Ninth Revision, (ICD-9) codes and Diagnostic-related Group (DRG) weights. I use these codes to construct 29 Elixhauser comorbidity dummies and Charlson comorbidity indices. I also observe the identity of the admitting service (e.g., “cardiology team 1”), within each of which patients are admitted for similar reasons (e.g., heart failure).

Housestaff schedules supplement my data on hospital admissions. I observe the identities and dates of each rotation block in residency training for housestaff in the same-hospital internal medicine residency program. This includes clinical rotations outside of the hospital (e.g., the affiliated community hospital), ICU and outpatient rotations, and research electives. Therefore, at any given time for any housestaff, I know the entire history of training, at least as summarized by rotation identifiers.

A key institutional fact qualitatively described above is that housestaff do not choose most of their learning experiences, at least in terms of their clinical rotations and in what order, peers and supervising physicians, and patients seen on the wards. Tables 2.4.1 and 2.4.2 show that interns and residents, respectively, with high- or low-spending practice styles are exposed to similar types of patients. The same tables also show that housestaff with high- or low-spending practice styles are similarly likely to be assigned to coworkers and attendings with high- or low-spending practice styles. Finally, because I find that rotations out to the affiliated community hospital have a substantial effect on intern test spending, I compare interns with different practice styles who do not have a community hospital rotation early in the year in Table 2.4.1 and find that they are similarly likely to have a community hospital hospital rotation later in the year. In the Appendix, I present more formal analyses on the exogenous assignment of housestaff physicians; I cannot reject the null that

¹⁰Results in this paper are robust to this censoring.

housestaff identities are jointly unrelated to patients types or other training experiences.

Finally, I have detailed information collected for each housestaff physician during the time of his residency application prior to graduating from medical school. This includes housestaff demographics, medical school, USMLE test scores, membership in the Alpha Omega Alpha (AOA) medical honors society, other degrees, and position on the residency rank list. USMLE test scores represent an objective, standardized measure of resident knowledge and ability. Position on the residency rank list represents desirability to the residency program; it could be made on a number of different criteria, including objective ability and promise to be a “leader” in a field of medicine. In addition to pre-residency characteristics, I also observe the track of each housestaff physician, for example whether he is a preliminary or categorical intern, or whether he is from another residency program, as choices committed to prior to starting residency.

2.5 Practice Style Distribution and Dynamics

In this section, I present my main analysis of the distribution of housestaff practice styles and its dynamics over time. There are two parts to this analysis, roughly corresponding to moments in two dimensions: variance across residents and covariance across time.

In the first dimension, I am interested in the variation across housestaff physicians at the same relative training time. Although previous researchers have predicted that physicians practicing together should converge in their practice styles, either by learning or by community norms, they have not shown significant convergence. However, their study designs preclude precise measures of practice styles within short intervals of time and observe physicians only after they have completed training. In this study, observing daily spending linked to physicians in training allows for precisely measuring practice styles within short, well-defined time intervals relative to the start of learning. Additionally, I am particularly interested in the change in variation when interns become residents, because this allows me to isolate the contribution of authority to practice-style variation. A unique feature of this setting is that each housestaff discontinuously gains authority upon switching from an intern to a resident, even though he is no wiser immediately after this change.

In the second dimension, I am interested in the covariance between practice styles of the same housestaff physician across time. Variance across physicians that is constant over time could reflect each physician converging to his own practice style or the lack of any convergence at all. The serial correlation of practice styles within physician across time provides another useful moment to separate learning from other mechanisms of intrinsic heterogeneity or authority. While all three mechanisms affect variance across physicians, only learning should affect covariance within physician across time. Specifically, if physicians have not completed their learning, then there should be low correlation among their practice styles across time. If physicians have indeed settled on their practice styles and are no longer updating their beliefs, then the correlation should approach 1 even if authority is changing. As with the first dimension, this covariance can be estimated at each point in time, in

order to evaluate the practice-style stability as a dynamic process.

2.5.1 Variation across Physicians

For a patient being treated on day t of admission a by intern i , resident j , and attending k , I specify log daily test costs as

$$Y_{aijkt} = \beta X_a + \eta_t + \xi_{it} + \xi_{jt} + \zeta_k + \nu_{aij} + \varepsilon_{aijkt}. \quad (2.5.1)$$

Equation (2.5.1) includes a rich set of patient and admissions characteristics X_a for admission a described in Section 2.4 and the sum of a set of time fixed effects that include the month-year combination, day of the week, and day of service relative to the admission day. I allow for attending fixed effects, ζ_k . To first order in the data, and consistent with the existing literature, the practice styles of attendings are relatively fixed. Further, attending physicians are not of interest in this analysis, and unlike housestaff physicians, they are not randomly assigned patients.¹¹

The parameters of interest in Equation (2.8) are the time-varying practice styles, ξ_{it} and ξ_{jt} for intern i and resident j , respectively, at time t . Recall that I observe the same housestaff as both interns and residents and that I am interested in the distribution of practice styles as a function of tenure, or the time relative to the start of residency. I set up the following notation to reflect this precisely: Denote by $r(h, t) \in \{I, R\}$ as the function that reports the role of housestaff h at time t , which takes the value I if housestaff h is an intern at time t and the value R if he is a resident. Define $\tilde{\tau}(h, t)$ as the exact tenure of housestaff h at time t ; $r(h, t) = R$ if and only if $\tilde{\tau}(h, t) \geq \tilde{\tau}^*$, where $\tilde{\tau}^*$ is the tenure at which interns become residents.

I impose that housestaff practice styles are fixed within relatively short intervals of tenure. That is, within a tenure interval τ , $\xi_{ht} = \xi_h^\tau$ for all t such that $\tilde{\tau}(h, t) \in \tau$. Otherwise I flexibly allow for any set of practice styles $\{\xi_h^\tau\}_{\tau \in T}$ across intervals. Given that housestaff are exogenously assigned patients, attendings, and peers, I specify their practice styles as random effects. Random effects have the additional advantage of addressing measurement error in practice styles: Although I observe relatively frequent daily test spending for each housestaff, under measurement error, fewer observations for a group of housestaff in a short interval of time will naturally drive up the estimated variance of a distribution of fixed effects. Random effects take this into account by fitting maximum likelihood estimates of standard deviations of ξ_{it} and ξ_{jt} as parameters in themselves. I allow ξ_{it} and ξ_{jt} as crossed random effects, which reflects the fact that I observe intern i working with residents other than j and resident j working with interns other than i . Finally, I allow for shocks at the admission-intern-resident level, ν_{aij} . This reflects that, even controlling for patient observables, some patients will naturally result in more test costs than others. This specification is more flexible

¹¹ Although I control for the admitting service, such as “cardiology team 1,” there can still be different types of attendings who on a specific service. It is possible to identify the types of attendings individually, and conditional on these types, I would expect closer to random assignment of patients to attendings. However, given that I focus on the practice styles of housestaff in this study, I defer this work for later.

than simply an admission-level shock ν_a but is computationally less burdensome with respect to the maximum-likelihood covariance matrix.

Table 2.5.1 and Figure 2.5.1 present results for the estimated standard deviations of the distributions of housestaff practice styles within each tenure interval τ . In my baseline specification, I consider non-overlapping tenure intervals that are 60-days in length. In addition to the estimated standard deviations of intern and resident practice styles, I also present the estimated standard deviation for the admission-intern-resident-level shock in Table 2.5.1. This gives a useful benchmark of unobserved variation at the patient (admission) level relative to the contribution of housestaff in terms of daily test spending.¹²

I find large and significant variation in housestaff practice styles during all intervals of time. An intern that has a practice style one standard deviation above the mean increases spending by about 15-30% on daily laboratory and radiology tests. A resident that has a practice style one standard deviation above the mean increases spending by about 55-80%. In comparison, the standard deviation for patient-housestaff-level shocks is 40%. Given the large qualitative heterogeneity across patients that characterizes care across ward services, I expect a large standard deviation for this distribution, and it is surprising that residents alone are responsible for more variation in spending than unobserved patient characteristics. Given that the mean test spending for an admission is \$714, this level of practice-style variation represents increases of approximately \$70 and \$500 for each standard deviation increase in intern and resident practice style, respectively.

There is little evidence of convergence in practice styles across housestaff within roles over time. The lack of convergence can reflect either the lack of learning or concurrent increases in authority interacted with intrinsic heterogeneity. It is also possible that the existence of convergence could also reflect *decreases* in authority, as there is no reason for authority to be monotonically increasing.¹³ We will address these issues further below. However, the striking finding in the variation of practice styles across time intervals is its sudden increase, at least doubling in standard deviation as interns become residents after one year of training. This suggests that authority determines much of practice-style variation. There should be no change in housestaff knowledge across this discontinuity, and indeed formal responsibilities of substantive patient care do not change either, but housestaff suddenly have more say when they become the “supervising resident” on the team as opposed to the intern, who by construction always has at least one year less experience.

2.5.2 Serial Correlation across Time

In the second part of the analysis, I study the serial correlation across estimated practice styles within housestaff across time intervals. I estimate housestaff practice styles in Equation (2.5.1) as

¹²In some sense, it provides an overestimate of the patient-level variation because it is actually a shock by the patient interacted with the intern and resident, the distribution of which could be wider. Although most patients are taken care of by only one intern and resident, some patients will be taken care of by more than one intern or resident.

¹³In fact, senior residents are known to be more “hands off,” in part because they are more comfortable with letting interns make their own decisions.

distributions of random effects, but I can estimate an empirical Bayes prediction for the practice style of each housestaff h during tenure interval τ . Specifically, this is done in a standard manner explained in Searle et al. (1992): First, the random effects model parameters are estimated by maximum likelihood. This includes both coefficient estimates (i.e., $\hat{\beta}$ and $\hat{\eta}$) and parameter estimates of the variance-covariance matrix (i.e., the standard deviation estimates $\hat{\sigma}_{\xi}^{\tau}$, $\hat{\sigma}_{\nu}$, and $\hat{\sigma}_{\epsilon}$). Overall random error terms are imputed by the data and coefficient estimates, and predicted random effects are shrunk by the variance-covariance estimates. The Bayesian shrinkage factor reconciles the fact that in finite samples, the distribution of fixed effects should be wider than the distribution parameter estimated for variance-covariance matrix would imply.

Using these empirical Bayes practice styles specific to housestaff h and tenure interval τ , $\tilde{\xi}_h^{\tau}$, I then estimate the serial correlation across intervals. That is, I measure $\text{corr}(\tilde{\xi}_h^{\tau}, \tilde{\xi}_h^{\tau-j})$, where j is the number of tenure intervals that separate practice style estimates for the same housestaff.¹⁴ In Figure 2.5.2, I show the estimated correlation coefficient between each tenure interval and the previous interval. In Table 2.5.2, I show corresponding coefficients not only between adjacent intervals but between intervals up to three intervals apart. Recall that each interval is composed of 60 days, so that comparing adjacent intervals is equivalent to comparing practice styles at average tenures 2 months apart.

In the first year of training, interns start off with practice styles that are almost completely unstable. The correlation between practice styles in the first two months of training and those in the next two months is not statistically different from 0. However, the serial correlation between practice styles increases with tenure. At the beginning of their second year, when interns become residents, the correlation coefficient with the previous interval is greater but still relatively small at 0.15. It then rapidly increases throughout the second year up to 0.80. In the third year, when the volume of patients seen decreases relative to the previous year, the serial correlation between intervals starts to decrease.

The serial correlation in practice styles estimated across tenure intervals suggest that practice styles become more stable with more experience. It also appears that experience as a resident has greater influence in shaping practice styles, or at least in developing stable ones, than does experience as an intern. The serial correlation of practice styles increases slowly during internship but does so much more quickly during residency. By the end of the second year, practice styles are remarkably stable. However, stability is not monotonic with training. In the third year, the decrease in serial correlation suggests that practice styles, at least during residency training, are not permanent. Even after practice styles have achieved a remarkable degree of stability, they can still become less stable.

¹⁴In addition to this simple calculation of the correlation coefficient, I calculate an alternative measure that adjusts for the number of observations used to calculate the correlation between a given pair of estimated random effects. Using assumptions implicit in Equation (2.5.1), I adjust the correlation coefficient downward for fewer observations, given the estimated variances of ν_{aij} and ϵ_{aijkt} . However, consistent with sufficiently many observations, this exercise does not noticeably change results.

The stability of practice styles is not only interesting in itself but also reflects upon the more fundamental mechanism of learning. The lack of convergence in practice styles across housestaff with time does not rule out learning, as housestaff physicians can both become more sure of what their practice style is yet also gain authority that would increase the practice-style variance across physicians. However, the increasing stability of practice styles is direct evidence of learning. Thus, it appears that housestaff do learn from the beginning of training and in fact learn at an even faster pace when they become residents with greater authority. Given that stability is not monotonically increasing, it also appears that the practice styles can be forgotten with less practice; I examine this in the next subsection.

2.5.3 The Role of Continued Practice

In order to test the hypothesis that stability requires continued practice, I estimate correlation coefficients for month-long tenure intervals that are separated by a month in between and condition on a nearby tenure interval having few or no patient observations. I consider two types of such nearby intervals, for a random effect pair $\tilde{\xi}_h^\tau$ and $\tilde{\xi}_h^{\tau-2}$: the interval $\tau - 1$ in between the pair, and the interval $\tau - 3$ prior to the pair.¹⁵ The reason for this conditioning is compare practice styles between the same tenure intervals and surrounded by the same relative frequency of patient encounters. If learning is entirely cumulative and does not depend on the continuity of practice, then the relative timing of the nearby tenure interval with few or no patient observations should have little effect; in fact, having many patients in the intervening interval $\tau - 1$ may reduce the correlation between the pair. However, if learned practice styles need reinforcement through continued practice for stability, then the opposite result should obtain: Continued practice during $\tau - 1$ should increase the correlation between the pair.

In Figure 2.5.3, I plot the correlation coefficients between random-effect pairs $\tilde{\xi}_h^\tau$ and $\tilde{\xi}_h^{\tau-2}$, where I define each interval τ as lasting a month.¹⁶ A lack of patients in the interval between those corresponding to the random effect pair is associated with a delay, by exactly a month, in the rise in stability after interns become residents. During the later two years of training, correlation also appears to be slightly lower for pairs with few or no patients in between (in interval $\tau - 1$ as opposed to interval $\tau - 3$). This evidence suggests that continued practice is important for both the initial stabilization of practice styles, when housestaff gain greater authority, and for the continued stability.

¹⁵I also consider the interval $\tau + 1$ after the pair, instead of $\tau - 3$ prior to the pair, and results are qualitatively similar.

¹⁶For the results in Figure 2.5.3, I define “few” patients as 40 patients or fewer, corresponding to the 20th percentile of monthly patient volume conditional on seeing any patients at the main hospital. I omit correlation coefficients calculated with fewer than 10 observations. Results are largely similar using different thresholds for defining few patients, and higher thresholds of observations for calculating correlation coefficients.

2.6 Observable Learning Experiences

The previous section shows that housestaff physicians each do appear to learn, or converge to, an individual practice style during training but that their practice styles ultimately are different across each other, in large part due to increasing authority. The next question, addressed in this section, is whether differences in their practice styles can be traced to observable differences in learning experiences.

Although a wide variety of experiences are in principle observable, in practice, given the sheer number of experiences accumulated during residency, it is difficult to summarize experiences into a reasonable number of meaningful measures. For example, Choudhry and colleagues (2006) have shown that physicians with patients who experienced a bleeding adverse event after being prescribed warfarin, a drug for atrial fibrillation known to have bleeding risks, reduced their use of warfarin for other patients, even those who should appropriately be prescribed it. Although this identifies the effect of an idiosyncratic event on one particular aspect of clinical practice, it is much more challenging to aggregate individual learning experiences into a measure that would affect overall spending. Furthermore, under the same training environments and with a large number of experiences, one would expect aggregate experiences to be roughly similar. Of course, experiences that are more salient in physicians' memories are likely to be more influential, but again it is difficult to determine which experiences are more salient for whom.

As rough measures, I summarize three different types of cumulative experiences at each point during internship: time spent working with high- or low-spending physicians and supervising residents, time spent on various rotations within the hospital (e.g., subspecialty, general medicine, ICU) as well as non-inpatient rotations (e.g., research, outpatient), and time spent on the inpatient rotation at the affiliated community hospital. Although housestaff have very similar experiences by the end of training, the order of these experiences, particularly rotations, mechanically differ in the beginning of training. That is, even though almost all housestaff spend at least one month at the community hospital, for many interns, this may not occur until later in the year.

There is reason to believe that experiences at the community hospital should have a greater effect on practice styles than experiences within the same teaching hospital. While experiences within the same institution share common resources and norms, experiences across institutions, between the main academic hospital and the community hospital, differ greatly in both of these. As discussed in Section 2.3, this is particularly true with respect to test costs, since patients at the community hospital are less complicated, attending physicians are uniformly less academic, the staff is less responsive or knowledgeable in performing uncommon tests, and resources for performing costly tests are more constrained. Finally, being exposed to different opinions within the same hospital may encourage housestaff to listen to none of them, as the diversity of opinions could only highlight the fact that there are many acceptable approaches. On the other hand, being exposed to a consistent, alternative way of doing things at the community hospital could result in a more predictable effect.

This phenomenon has been explored theoretically, for example, by Ellison and Fudenberg (1993).

2.6.1 Effect of the Community Hospital

To evaluate the effect of the community hospital, I primarily conduct an event study in which I measure the effect on test spending for interns during various interval before and after they have spent time at the community hospital. Virtually all categorical same-hospital housestaff spend time at the community hospital at some point in their internship, and so the primary identifying variation lies in the timing at which they have their community-hospital rotation.

Denote $\tilde{\tau}_c(i, t)$ as the time relative to the last day at the academic hospital prior to going to the community hospital or relative to the first day at the academic hospital after returning from the last day at the community hospital. If $\tilde{\tau}_c(i, t) < 0$, then t is prior to any experience at the community hospital; if $\tilde{\tau}_c(i, t) > 0$, then t is after all experience at the community hospital during intern year. Considering month-long intervals τ_c , I then estimate the following regression of log daily test costs:

$$Y_{aijkt} = \sum_{\tau_c \in T_c} \alpha_{\tau_c} 1(\tilde{\tau}_c(i, t) \in \tau_c) + \beta X_a + \gamma \tau(i, t) + \eta_t + \zeta_i + \zeta_j + \zeta_k + \varepsilon_{aijkt}, \quad (2.6.1)$$

for admission a , intern i , resident j , attending k , and day t . The coefficients of interest are α_{τ_c} , which correspond to the effect of the community hospital at each time interval τ_c , before and after the community experience. In this model, I assume fixed effects for housestaff (ζ_i and ζ_j) as well as attendings (ζ_k), as I am not interested in modeling the dynamics of the entire unobserved housestaff practice styles. I also control for intern i 's tenure at time t , $\tau(i, t)$.

Figure 2.6.1 shows coefficients for the event study at each month for five months before and after rotations at the community hospital, as well as for times before and after five months. Observations at the main hospital in between rotations at the community hospital, for interns who had non-contiguous rotations there, are normalized with coefficient of 0. Coefficients for time intervals are relatively constant within the periods before and after the community hospital, but moving across periods reveals a large decline in costs. Figure 2.6.1 shows the effect in terms of log daily test costs, which is my preferred outcome measure, as test costs are right-skewed. In the Appendix, I show similar results for the event study using untransformed daily tests costs; these results suggest that the community hospital decreases daily test spending by about \$200 on average, greater than one standard deviation in the daily test cost distribution. Coefficient confidence intervals also narrow immediately after time spent at the community hospital; the number of observations before and after the community hospital are mechanically the same, as rotations to the community hospital are spaced evenly throughout the year.

In addition to the event study, I perform a number of simple difference-in-difference specifications,

$$Y_{aijkt} = \alpha 1(\tilde{\tau}_c(i, t) > 0) + \beta X_a + \gamma Z_{it} + \eta_t + \zeta_i + \zeta_j + \zeta_k + \varepsilon_{aijkt}, \quad (2.6.2)$$

using a number of different controls X_a for the admission a and Z_{it} for intern i 's training history at time t . Presented in Table 2.6.1, results are robust under the wide variety of controls, as long as intern tenure is controlled for. Consistent with the event study, these results again suggest that the community hospital roughly decreases daily spending by about \$200. Since the effect is large, it is difficult to quantify in percentage terms using log test costs as the outcome, but it appears that test costs are reduced by at least half.

Finally, I explore heterogeneous treatment effects of the community hospital, depending on whether interns are assigned to rotations earlier in the year or later in the year. I define whether each intern is in the “earlier” or “later” subgroup based on his average tenure while at the community hospital during intern year. I then estimate an event study for each intern subgroup, using Equation (2.6.1) and controlling for intern tenure. Shown in Figure 2.6.2, the community hospital rotation has a substantially greater effect on the test spending of interns with earlier rotations versus those with later rotations.¹⁷

These results suggest that experience at the community hospital results in a large and relatively permanent effect on intern spending, reducing daily test spending by more than half or by about \$200. Given that interns do not have much variation in practice styles, this represents a significant source of the variation.¹⁸ In addition, improvement in the precision of practice styles after the community hospital also suggests that this common experience reduces variability at the same time as it lowers mean spending. Finally, the community hospital has a greater effect on interns with less cumulative experience at the time of their rotation there. Taken together, this evidence suggests that experiences across institutional borders, especially earlier experiences, have the potential to have shape practice styles in a significant and lasting way. Learning in this case may occur not only in the classical sense of gaining information but also by exposure to different institutional norms.

2.6.2 Effect of Other Observable Learning Experiences

Housestaff have a wide range of other rotations throughout their training. In some sense, these rotations can be potentially more varied than the difference between the community hospital and main academic hospital. For example, in the outpatient rotation, housestaff do not see patients who are sick enough to be admitted to the hospital but rather focus on non-life-threatening or chronic medical issues. Even more differently, housestaff on research elective do not see any patients. Within the same hospital, there are subspecialty or ICU rotations that entail care of complicated patients

¹⁷I find that whether an intern has an earlier or later community hospital rotation does not explain pre-community-hospital spending, conditional on intern tenure but not controlling for intern identity as I do in Equation (2.6.1). This is consistent with the institutional detail that interns are not assigned rotations with their characteristics in mind. Of course, since I control for intern identities (interacted with the identities of the resident and attending), selection based on fixed characteristics is not a threat to identification.

¹⁸Because virtually all interns spend time at the community hospital, and because the effect is mostly stable (does not increase or decrease within intern year), I cannot study its effect on practice styles into the second and third years of residency.

with illnesses such as advanced heart failure or leukemia blast crises who need management by subspecialists. It is possible that these rotations could also have an effect on general practice styles, or rather, experiences on these rotations may be less transferrable because they are so different.

I examine a wide range of regressions of the form

$$Y_{aijkt} = \sum_{m \in M} \alpha_m \text{RotationHx}_{it}^m + \beta X_a + \eta_t + \zeta_i + \zeta_j + \zeta_k + \varepsilon_{aijkt}, \quad (2.6.3)$$

where RotationHx_{it}^m represents the cumulative experience in rotation type m for intern i at time t , for m in some subset M . I find that the coefficients of interest, α_m , are all small and statistically insignificant. That is, differences in experience to date in these rotations other than the community hospital do not have a noticeable effect on housestaff practice styles.

Next, within rotations, interns may have different experiences with the patients they care for or the attendings and supervising residents they are assigned to. Again, although they are expected to have the same experiences within rotation, the idiosyncratic nature of inpatient medicine may mean that they happen to work during a week that is busier than usual or be assigned to a supervising physician that has a high- or low-spending practice style. The more experience interns accumulate, the less these observable idiosyncratic experiences should matter. It could be possible for some behavioral reason that learning occurs by a few, potentially extreme experiences, but theory predicts that exposure to a wide variety of experiences should have less influence than exposure to a narrow and consistent norm (Ellison and Fudenberg, 1993), and psychology has shown that people have a greater preference for middle choices in a given context than the extremes (Kamenica, 2008).

In order to identify high- or low-spending supervising physicians, I first estimate fixed effects, potentially also adjusting with a Bayesian shrinkage factor, for these physicians. I then construct measures of experience with different types of attendings and supervising residents, with either high- or low-spending practice styles, $\text{PhysicianHx}_{it}^n$, for each intern i 's history of working with physician type n up to time t . For experiences with attending physicians, I consider both attendings whom intern i has worked with in the same field as the current admission a and all attendings the intern has worked with, regardless of the field.

I then perform regressions similar in form to (2.6.3):

$$Y_{aijkt} = \sum_{n \in N} \alpha_n \text{PhysicianHx}_{it}^n + \beta X_a + \eta_t + \zeta_i + \zeta_j + \zeta_k + \varepsilon_{aijkt}, \quad (2.6.4)$$

with coefficients of interest α_n , again for n in some subset of physician history types N . These coefficients are again small and statistically insignificant over the entire range of experiences with attendings and residents.

Finally, I examine the effect of the history of patient censuses and types on housestaff practice styles. Because new patients are seen frequently, this has the least chance of yielding differences in learning experiences. Again, housestaff physicians may remember certain patients they treated long

after the fact, but it is unclear which patients will be more memorable or influential. Regressions similar in form to Equations (2.6.3) and (2.6.4) but with analogous measures of experiences with patients show no significant effect on practice styles.

2.7 Intrinsic Heterogeneity

In this section, I examine evidence for intrinsic heterogeneity as a reason for practice-style variation. The literature has found that rough physician characteristics explain very little of practice-style variation. However, it is difficult to control for different learning experiences when comparing the effect of physician characteristics in the cross-section. This setting provides an advantage of housestaff undergoing the same training but starting residency with different characteristics. Of course, the question is not whether being male causes one to spend more or less. Rather, because meaningful intrinsic heterogeneity is often difficult to observe, the analysis starts with the assumption that housestaff characteristics are correlated with intrinsic heterogeneity, reflecting different preferences or abilities, that could then cause different practice styles to emerge.

In addition to the usual rough demographic and training history, such as age, sex, ethnicity, and medical school, this institutional setting affords a richer set of characteristics that are potentially more revealing about intrinsic heterogeneity. For example, a physician's choice of career reveals a lot about his preferences for income, prestige, and work lifestyle. However, of course, studies do not usually observe radiologists working as internists, or vice versa, partly because this is impractical or infeasible after residency.¹⁹ During residency, future radiologists, dermatologists, anesthesiologists, psychiatrists, and ophthalmologists are all observed with the same prior training and performing the same work. This yields a unique opportunity to study the effect of meaningful heterogeneous preferences on practice styles.

Also, I am able to describe housestaff with detailed information collected and generated by the residency program during its selection process. First, I observe rich markers of ability, such as USMLE test scores and membership in the AOA medical honor society. Second, I observe the desirability of each housestaff relative to his peers through his position on the rank list. Residency placement occurs through the National Resident Matching Program (NRMP), by which medical students and residency programs jointly rank each other and are prohibited from revealing their ranks of each other prior to being matched by an algorithm. Therefore, positions on the residency rank list are true measures of desirability and reflect no additional considerations such as strategic communication or the likelihood that a housestaff will actually choose that program. However, because the residency program I study is highly competitive, differences among housestaff in their position on the rank list are usually minor.

¹⁹Board-certified radiologists can actually practice as general practitioners, given that they have completed internships in internal medicine, which is all that is required by law. Of course, this does not happen because radiologists have revealed to prefer the income and lifestyle of radiology and have already invested in training. Additionally, they would be out of practice in medicine after spending most of their training in radiology.

Furthermore, desirability is a multidimensional concept that considers ability, interests, and subjective personality traits. Different types of future physicians are indeed recruited and judged within type. This critique holds for the rank list and to some degree to AOA membership. But the USMLE provides a measure of one dimension, that of medical knowledge and reasoning, that might be particularly relevant to diagnostic testing.²⁰ The USMLE is not only standardized nationally, it is also a difficult test on which receiving a perfect score is virtually unheard of. It therefore provides a good sense of clinical ability even at its higher tail.

Finally, I have the opportunity to study housestaff from other residency programs and another academic hospital rotating onto the internal medicine ward services at the main hospital. As housestaff begin training with close to no clinical experience, I can observe interns from other programs and ask whether they have different practice styles that might more purely reflect intrinsic heterogeneity. Moreover, I can also observe housestaff as residents from the other hospital, when they have accumulated different experiences and learned different institutional norms, in addition to the initial heterogeneity, and examine differences in practice styles then.

2.7.1 Characteristics and Track Choices Prior to Residency

I perform the following regression for each housestaff characteristic, including residency track choice, prior to starting residency:

$$Y_{aijkt} = \alpha_m HousestaffGroup_h^m + \beta X_a + \eta_t + \zeta_{-h} + \zeta_k + \varepsilon_{aijkt}, \quad (2.7.1)$$

where $HousestaffGroup_i^m$ is an indicator variable that takes the value of 1 if housestaff h had characteristic (or made track choice) m prior to starting residency. The coefficient of interest is α_m , which represents the effect that characteristic m has on daily test spending. I include fixed effects for the other housestaff on the team, ζ_{-h} , which would be the intern if the index housestaff is the resident or the resident if the index housestaff is the intern.

I consider a wide variety of characteristics, including demographics (age, sex, race), whether the housestaff came from a “rare” medical school (e.g., a less prestigious medical school that usually does not send graduates to this competitive residency program or from a foreign medical school), USMLE test scores, other degrees prior to starting residency (e.g., J.D., Ph.D, or M.B.A.), and position on the rank list. Of these characteristics, the only measure that would be expected *a priori* to be related to clinical reasoning and skill would be the USMLE test score.²¹ Other characteristics, however,

²⁰I specifically use the USMLE Step 1 test, which is taken after the second year of medical school and is most important for applications to residency. As a result, the effort devoted to preparing for this test is uniformly high for all medical students. The USMLE Step 2 test is often absent in applications and is viewed as both less difficult and less important.

²¹As described above, position on the residency rank list reflects multidimensional considerations and therefore should be less related to clinical skill. Further, the admissions committee’s assessment of clinical skill is mostly based on the USMLE test score with some additional consideration on recommendation letters.

could be correlated with clinical skill. For example, given an overall threshold of desirability, having a Ph.D. with an impressive research portfolio might actually be correlated with lower clinical skill. In addition, I consider choices of residency track, including preliminary versus categorical, and primary care versus categorical. Recall that the preliminary-track interns plan to continue with another residency program after their internship; primary care housestaff broadly have the same three-year training plan in internal medicine, but they plan to emphasize slightly more outpatient care and have signaled that they are unlikely to subspecialize.

In part of Table 2.7.1, I show a subset of results for Equation (2.7.1). Effects of pre-residency characteristics and track choices are generally small and insignificant. This notably includes preliminary (future radiologists and dermatologists) versus categorical interns (future internists or medicine subspecialists), as well as housestaff high on the rank list versus others. There are two characteristics that predict statistically significant lower spending: male sex and high USMLE test score. Male interns have 2% lower daily spending costs, significant at the 10% level; male residents have 4% lower daily spending costs, significant at the 5% level. A high USMLE score has no statistical effect on spending for interns but predicts 3% lower daily spending, significant at the 10% level, for residents.

These results show that intrinsic heterogeneity, to the extent that it is correlated with characteristics and pre-residency choices, may play some role in generating practice-style variation, but not likely a major role, compared to the large variation and relative instability in Section 2.5, and compared to the effect of the community hospital in Section 2.6. However, it is interesting that the effect of intrinsic heterogeneity increases with tenure. This could be further support for the idea that as authority grows, intrinsic heterogeneity is increasingly allowed to be expressed. Alternatively, it could suggest that learning – what is retained and what is used as a guide for future decisions – is influenced by intrinsic heterogeneity. These two ideas are not mutually exclusive.

2.7.2 Housestaff from Other Residency Programs

I also perform regressions of the form of Equation (2.7.1), in which for $HousestaffGroup_h^m$ I use indicators for other residency programs that housestaff h might belong in. Recall that interns from other residency programs within the same hospital (e.g., obstetrics-gynecology and emergency medicine) and both interns and residents from another academic hospital rotate onto the main hospital's internal medicine wards. I distinguish these regressions from the previous subsection because these housestaff groups represent not only choices and selection made prior to residency (therefore related to intrinsic heterogeneity) but also differences in learning experiences, since they spend most time training within their main residency program.

I present results for these regressions also in Table 2.7.1. I find that interns from other residency programs, either across departments in the same hospital or across hospitals in internal medicine, do not have significantly different mean spending practice styles. However, medicine residents from the other hospital spend 17% less than residents of the main hospital.

Again, it appears that practice styles diverge with more training. In the case of residents from the other hospital, this could reflect intrinsic heterogeneity, with greater authority between intern and resident roles, or different learning experiences. As mentioned in Section 2.3, the other hospital is known to be less research-oriented and less prestigious. However, perhaps its greatest difference from the perspective of housestaff training is that, because of its proximity to a hospital with high prestige, it has substantially fewer wealthy private donors and research endowments, as well as lower bargaining power with local insurance providers for reimbursement rates. As a result, this other hospital has consistently had greater pressure to reduce spending, for example by substituting inpatient test with outpatient follow-up or requiring housestaff to enter repeated laboratory tests each day at a time. Otherwise, the quality of residents between the two institutions is not known to be substantially different, and in fact the two hospitals are often ranked by the same medical students.

These institutional facts suggest that lower spending by residents from the other hospital may be more due to learning experiences, in particular being exposed to a different institutional norm, than by intrinsic heterogeneity. Taken together with the other evidence in this section, intrinsic heterogeneity appears to play a relatively minor role in influencing practice styles, at least on average effects correlated with observable characteristics, while learning plays a larger role.²²

2.8 Conclusion

This paper studies the dynamics of housestaff physician practice styles in order to shed light on the development of practice-style variation that has received much attention both in previous research and in policy discussions. Focusing on physicians in training at a single institution, I separate practice styles into components of learning, authority, and intrinsic heterogeneity at a time when physicians are just beginning to learn medicine.

I find significant variation in practice styles from the beginning of training, although practice styles are initially far from stable. As training proceeds, practice styles become more stable but also more varied. In particular, there is a sharp discontinuity when housestaff physicians change roles from intern to resident. Although there are no formal differences between the roles, the effect of having greater tenure than the other housestaff coworker at least doubles practice-style variation and leads to a steep rise in stability. These dynamics suggest that practice styles are learned, that much of their variation is the result of authority, and that learning accelerates with authority.

I also examine the effect of observable experiences within residency and characteristics determined prior to residency in order to find specific evidence of learning and intrinsic heterogeneity, respectively. Unsurprisingly, learning experiences with a meaningful effect on average practice styles

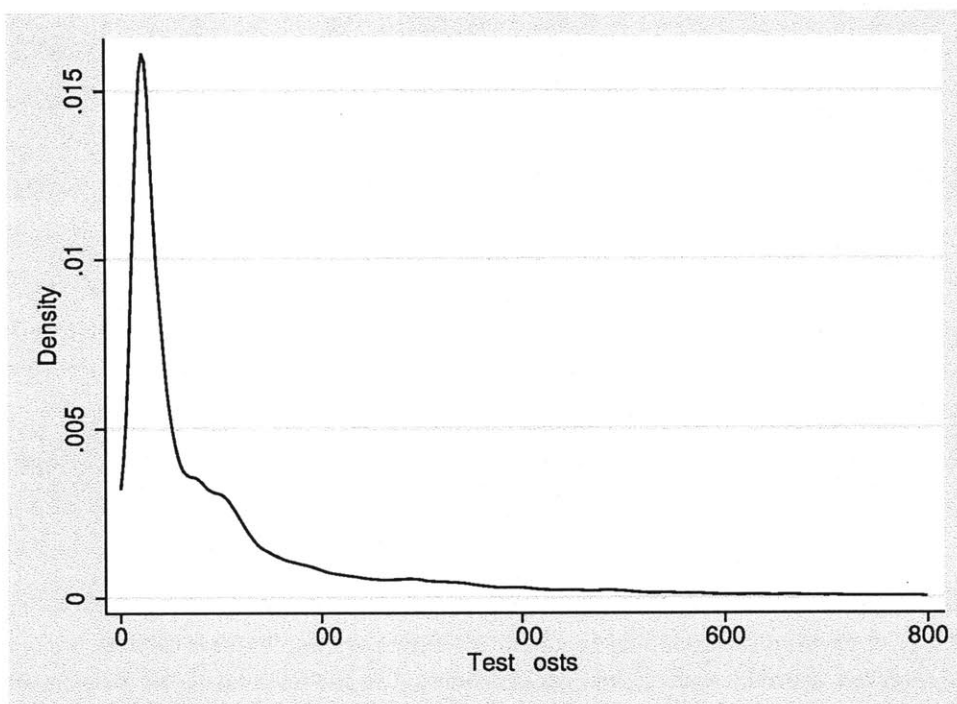
²²This intrinsic heterogeneity is of course conditional on selecting into having some experience at this academic medical center. Nonetheless, as described above, it includes housestaff from different training programs and with different career trajectories, in some ways reflecting greater intrinsic heterogeneity within the field of medicine across different hospitals.

are difficult to summarize even if they are observable. However, I find a significant effect of experience at an affiliated community hospital, reducing average intern spending by more than half. Similarly, compared to most categories of housestaff characteristics and choices, the most notable category spans institutions: Residents from another academic hospital with a culture for less spending incur 17% lower test costs. These findings suggest that institutions provide context for the strongest influences on practice styles and that their effect lasts beyond the institution.

The findings in this paper suggest that practice styles are far from predetermined by intrinsic physician characteristics, nor are they predictably determined by objective descriptions of training experiences, at least within institutional boundaries. They are dynamically formed in what appears to be an individualized learning process. This is not to say that large variation is inevitable. One perhaps trivial approach to reducing variation would be to limit physician authority. For example, expensive tests could require further justification or authorization prior to ordering. Second, behavior could be shaped by changing institutional norms. The institutional environment does not only have a direct mechanical effect of limiting what physicians do. Rather, this study suggests that habits persist even when rotating back to the original institution, and that habits may be more stable when learned under greater authority.

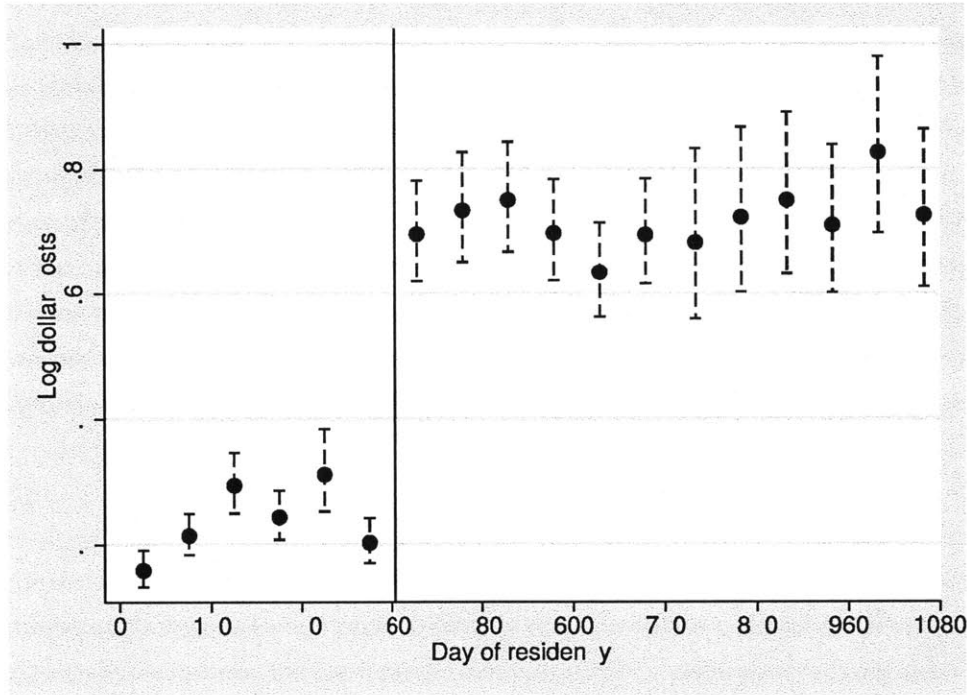
The choice of laboratory and radiology testing as the measure of practice styles also highlights that, in much of medicine, very little is known about the appropriate level of care. Although some treatment choices conditional on diagnosis are obvious, such as aspirin in myocardial infarction, much of diagnostic medicine remains an art, and the proliferation of testing technologies presents clinicians with choices on which there are usually no guidelines and virtually no evidence. This seems to suggest not only that practice-style variation may likely grow in the near future but also that a fundamental issue is for researchers, institutions, and policy-makers themselves to learn what is appropriate.

Figure 2.4.1: Distribution of Daily Test Spending



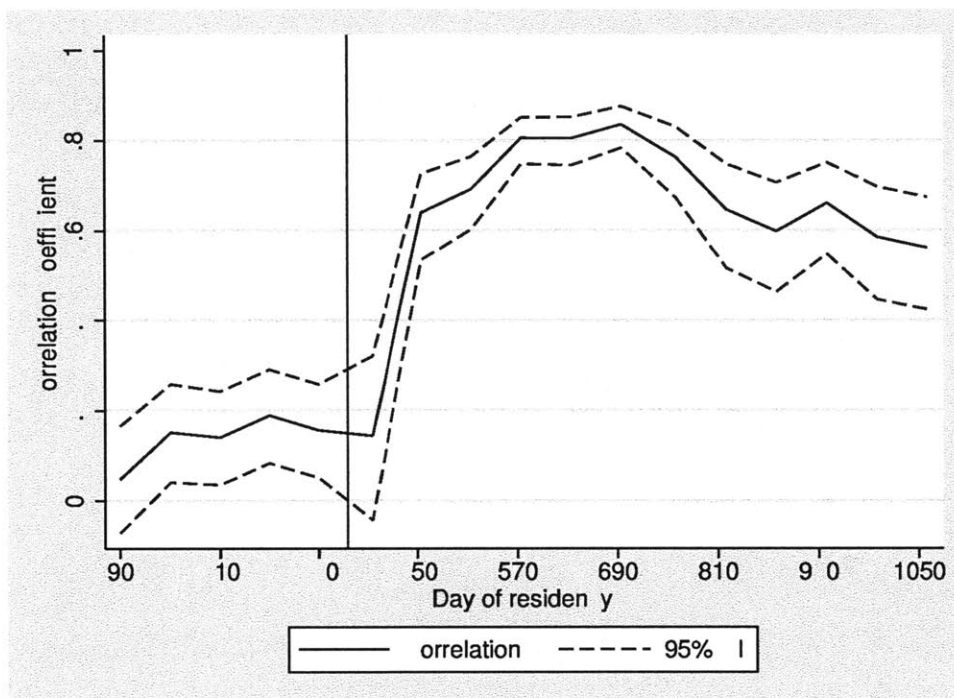
Note: This figure shows the density daily test costs. The distribution is shown up to \$800 per day.

Figure 2.5.1: Standard Deviation of Practice-style Random Effects across Housestaff Tenure



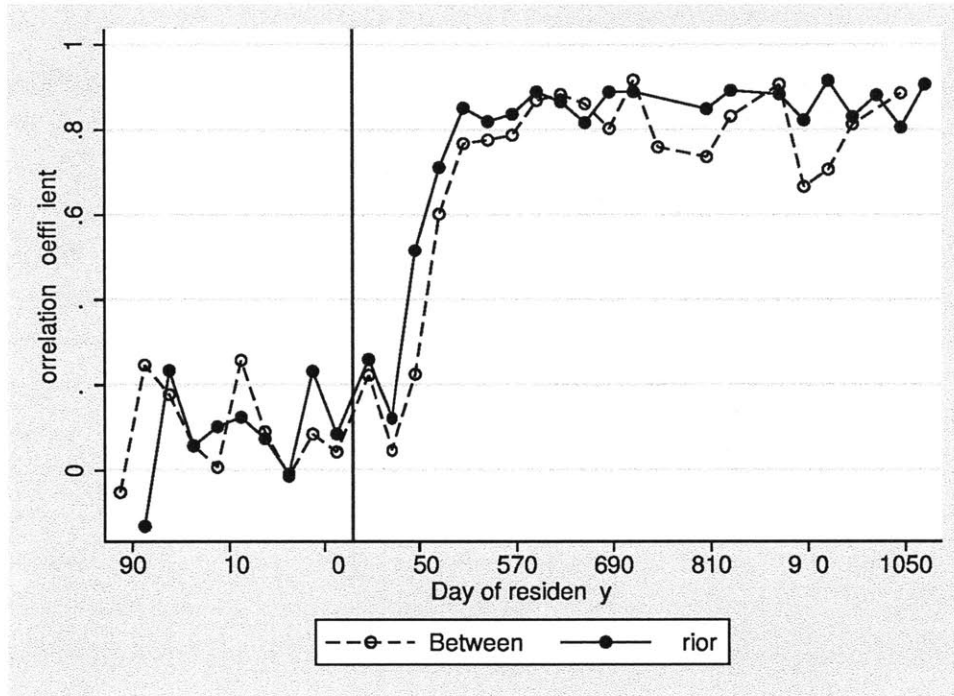
Note: This figure shows the standard deviation in a random effects model of log daily test costs shown in Equation (2.5.1) at each non-overlapping two-month tenure interval. Point estimates are shown as dots; 95% confidence intervals are shown as dashed capped lines. The model controls for patient and admission observable characteristics, time dummies (month-year interactions, day of the week), and attending identities (as fixed effects). Patient characteristics include demographics, Elixhauser indices, Charlson comorbidity scores, and DRG weights. Admission characteristics include the admitting service (e.g., “cardiology team 1”). Intern and resident practice styles are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. The standard deviation of housestaff practice-style distribution at each tenure interval is estimated by maximum likelihood. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical red line denotes the one-year tenure mark.

Figure 2.5.2: Serial Correlation of Practice-style Random Effects across Housestaff Tenure



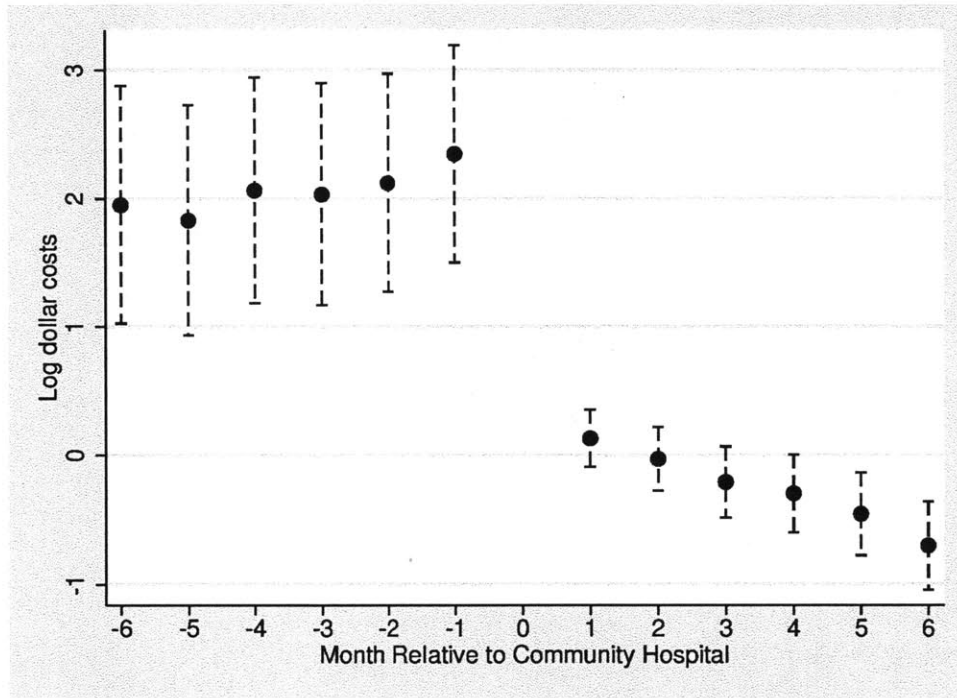
Note: This figure shows the serial correlation between the empirical Bayes random effect of each housestaff during a tenure interval and the random effect of that same housestaff during the previous tenure interval. Correlation coefficients are plotted as a solid line; 95% confidence intervals are plotted as dashed lines. The random effect model of log daily test costs is first estimated as in Equation (2.5.1), as described in the notes for Figure 2.5.1, using data within each two-month interval. Then the empirical Bayes random effects are calculated in the standard manner, described in Section 2.5.2, for each housestaff and tenure interval. Correlation coefficients are then calculated within housestaff and across adjacent tenure intervals. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical red line denotes the one-year tenure mark.

Figure 2.5.3: Serial Correlation of Practice-style Random Effects Conditional on a Nearby Interval with Few Patients



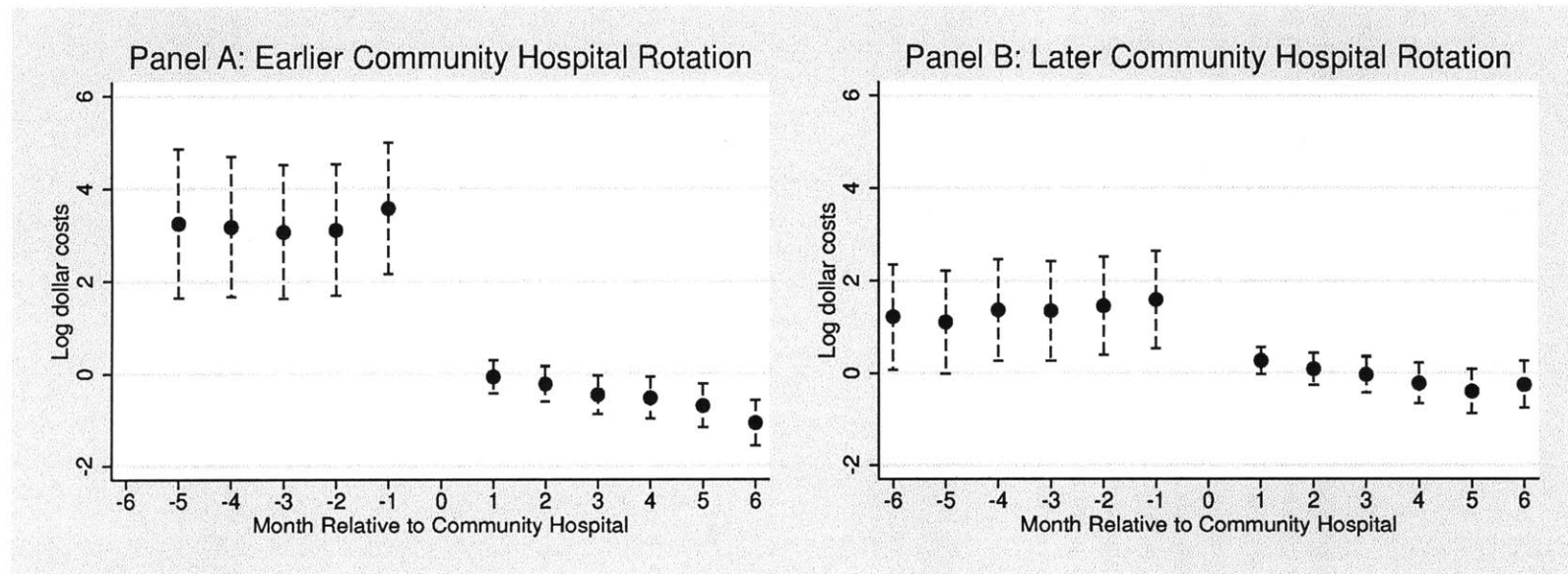
Note: This figure shows the serial correlation between empirical Bayes random effects in one-month intervals within housestaff, conditional on a nearby interval with few (40 or fewer, corresponding to the 20th percentile of monthly patient volume) or no patients. Random effects are calculated as described in Section 2.5.2, similar to Figures 2.5.1 and 2.5.2, except using data within one-month rather than two-month intervals. Correlation coefficients are calculated for a housestaff-month random effect and the random effect of the same housestaff corresponding to the month two months prior, i.e., between $\tilde{\xi}_h^\tau$ and $\tilde{\xi}_h^{\tau-2}$ in Equation (2.5.1). Correlations for random effects with two types of conditions are calculated: those conditional on few or no observations during the tenure interval $\tau - 1$ ("Between," shown with hollow blue dots), and those conditional on few or no observations during the tenure interval $\tau - 3$ ("Prior," shown with solid red dots). Correlations with fewer than 10 observations are omitted. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical red line denotes the one-year tenure mark.

Figure 2.6.1: Event Study of Effect of Community Hospital Rotation on Intern Log Spending



Note: This figure shows estimated coefficients in the event study, represented in Equation (2.6.1). The outcome variable is log daily test costs, and the coefficients of interest correspond to months relative to the intern starting the community hospital rotation (for negative numbers) or relative to ending the community hospital rotation (for positive numbers). Relative times greater than six months before starting the community hospital rotation or greater than six months after ending the community hospital rotation are included in months -6 and 6 , respectively. Month 0 includes any observations at the main hospital in between starting the first community rotation and ending the last rotation (for interns who have more than one rotation). The Equation controls for provider identities (including the intern) and for housestaff tenure. Standard errors are clustered by admission. Point estimates are shown as dots; 95% confidence intervals are shown as dashed capped lines.

Figure 2.6.2: Event Study of Effect of Community Hospital Rotation on Intern Spending, Earlier and Later Rotations



Note: This figure shows estimated coefficients in the event study, represented in Equation (2.6.1), for interns who were assigned earlier rotations at the community hospital (Panel A) and those who were assigned later rotations at the community hospital (Panel B). Each intern is categorized in one of these two subgroups based on his average tenure while at the community hospital during intern year. The outcome variable is log daily test costs in dollars, and the coefficients of interest correspond to months relative to the intern starting the community hospital rotation (for negative numbers) or relative to ending the community hospital rotation (for positive numbers). Relative times greater than six months before starting the community hospital rotation or greater than six months after ending the community hospital rotation are included in months -6 and 6 , respectively. Month 0 includes any observations at the main hospital in between starting the first community rotation and ending the last rotation (for interns who have more than one rotation). The Equation controls for provider identities (including the intern) and for housestaff tenure. Standard errors are clustered by admission. Point estimates are shown as dots; 95% confidence intervals are shown as dashed capped lines.

Table 2.4.1: Exogenous Assignment for Interns with Above or Below Average Spending

	Interns	
	Below-median test spending	Above-median test spending
<i>Patient characteristics</i>		
Age	62.11 (16.90)	62.13 (16.86)
Male	0.484 (0.500)	0.482 (0.500)
White race	0.706 (0.455)	0.703 (0.457)
Black race	0.161 (0.367)	0.159 (0.365)
Charlson comorbidity index	2.87 (2.79)	2.87 (2.79)
Diagnostic-related Group (DRG) weight	1.25 (0.86)	1.25 (0.84)
<i>Supervising physicians</i>		
Above-median-spending residents	0.500 (0.501)	0.500 (0.501)
Above-median-spending attending	0.503 (0.501)	0.502 (0.501)
<i>Rotations</i>		
Community hospital rotation (conditional hazard)	0.624 (0.487)	0.630 (0.485)

Note: This table shows evidence of exogenous assignment for interns with below-median or above-median averaged practice styles. Averaged practice styles are estimated by a regression of log test spending on patient characteristics and physician (intern, resident, and attending) identities. Lower- and higher-spending interns are identified by their fixed effect relative to the median fixed effect. For each of these groups of interns, this table shows average patient characteristics, practice styles for supervising physicians, and the conditional hazard of having a rotation at the community hospital later in the year. For patient characteristics and practice styles for supervising physicians, intern practice styles are estimated over the entire year. For the community hospital, practice styles are estimated over the first half of the year. Then among the sample of physicians who did not have a community hospital rotation in the first half of the year, the probability of having a rotation in the second half of the year is used as the conditional hazard. Averages are shown with standard deviations in parentheses.

Table 2.4.2: Exogenous Assignment for Residents with Above or Below Average Spending

	Residents	
	Below-median test spending	Above-median test spending
<i>Patient characteristics</i>		
Age	62.07 (16.82)	62.15 (16.93)
Male	0.489 (0.500)	0.478 (0.500)
White race	0.708 (0.455)	0.702 (0.457)
Black race	0.157 (0.364)	0.162 (0.368)
Charlson comorbidity index	2.84 (2.77)	2.90 (2.81)
Diagnostic-related Group (DRG) weight	1.27 (0.85)	1.24 (0.84)
<i>Supervising physicians</i>		
Above-median-spending attending	0.501 (0.501)	0.502 (0.501)

Note: This table shows evidence of exogenous assignment for residents with below-median or above-median averaged practice styles. Averaged practice styles are estimated by a regression of log test spending on patient characteristics and physician (intern, resident, and attending) identities, estimated over the entire data (including both years that the housestaff is a resident). Lower- and higher-spending residents are identified by their fixed effect relative to the median fixed effect. For each of these groups of residents, this table shows average patient characteristics and practice styles for attending physicians.

Table 2.5.1: Standard Deviation of Practice-style Random Effects across Housestaff Tenure

Days during year	Random effect standard deviation			
	Intern	Resident	Admission	Daily
1-60	0.159 (0.015)	0.697 (0.041)	0.398 (0.010)	1.174 (0.005)
61-120	0.214 (0.017)	0.653 (0.037)	0.427 (0.010)	1.182 (0.005)
121-180	0.294 (0.026)	0.550 (0.038)	0.450 (0.010)	1.184 (0.005)
181-240	0.244 (0.021)	0.576 (0.035)	0.442 (0.010)	1.178 (0.005)
241-300	0.311 (0.035)	0.494 (0.044)	0.432 (0.010)	1.159 (0.005)
301-365	0.203 (0.019)	0.627 (0.037)	0.422 (0.010)	1.174 (0.005)

Note: This table reports estimated standard deviations of random effect distributions. As stated in Equation (2.5.1), the model controls for patient and admission observable characteristics, time dummies (month-year interactions, day of the week), and attending identities (as fixed effects). Patient characteristics include demographics, Elixhauser indices, Charlson comorbidity scores, and DRG weights. Admission characteristics include the admitting service (e.g., “cardiology team 1”). Intern and resident practice styles are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. This model is estimated during 2-month time windows throughout the year.

Table 2.5.2: Serial Correlation of Practice-style Random Effects across Housestaff Tenure

Days during Training	Random effect correlation with previous period		
	One period prior	Two periods prior	Three periods prior
61-120	0.048 (-0.071, 0.166)		
121-180	0.151 (0.041, 0.258)	0.061 (-0.079, 0.199)	
181-240	0.140 (0.035, 0.242)	0.057 (-0.057, 0.170)	0.142 (-0.009, 0.286)
241-300	0.189 (0.083, 0.291)	0.203 (0.093, 0.308)	0.031 (-0.089, 0.151)
301-365	0.156 (0.051, 0.258)	0.071 (-0.042, 0.181)	0.147 (0.034, 0.256)
366-425	0.144 (-0.043, 0.321)	0.193 (0.006, 0.367)	0.124 (-0.066, 0.306)
426-485	0.640 (0.534, 0.727)	0.245 (0.076, 0.399)	0.179 (0.007, 0.341)
486-545	0.692 (0.603, 0.764)	0.395 (0.247, 0.525)	0.133 (-0.038, 0.297)
546-605	0.806 (0.748, 0.851)	0.763 (0.691, 0.820)	0.346 (0.195, 0.480)
606-665	0.805 (0.745, 0.852)	0.677 (0.583, 0.753)	0.552 (0.429, 0.656)
666-730	0.836 (0.784, 0.876)	0.782 (0.713, 0.836)	0.699 (0.605, 0.774)
731-790	0.763 (0.673, 0.831)	0.722 (0.617, 0.803)	0.753 (0.650, 0.829)
791-850	0.647 (0.516, 0.748)	0.601 (0.451, 0.718)	0.677 (0.541, 0.778)
851-910	0.598 (0.463, 0.706)	0.524 (0.367, 0.652)	0.438 (0.259, 0.589)
911-970	0.661 (0.547, 0.751)	0.430 (0.266, 0.569)	0.499 (0.342, 0.629)
971-1030	0.585 (0.446, 0.697)	0.616 (0.481, 0.723)	0.494 (0.332, 0.628)
1031-1095	0.561 (0.424, 0.674)	0.580 (0.446, 0.689)	0.538 (0.393, 0.656)

Note: This table reports correlation coefficients between random effects of the same housestaff across periods. As in Equation (2.5.1), the random effects model controls for patient and admission characteristics, time dummies (month-year interactions, day of the week), and attendings identities (as fixed effects). It includes random effects for housestaff and random shocks at the admission level. This model is estimated during each non-overlapping 2-month period. Next, empirical Bayes predictions for housestaff random effects are calculated for each housestaff observed in each time period. Finally, correlation coefficients are calculated between time periods within the same housestaff, including across the switch in roles from intern to resident. Correlation coefficients are displayed next to 95% confidence intervals in parentheses.

Table 2.6.1: Effect of Community Hospital Rotation on Intern Spending

	Log daily test costs			Daily test costs		
	(1)	(2)	(3)	(4)	(5)	(6)
Prior to first community hospital day	2.255*** (0.429)	2.385*** (0.438)	2.357*** (0.439)	187.74*** (57.78)	211.71*** (59.00)	203.92*** (59.10)
After last community hospital day	0.012 (0.111)	0.106 (0.114)	0.098 (0.114)	11.29 (14.99)	20.18 (15.41)	18.21 (15.40)
Intern tenure	Y	Y	Y	Y	Y	Y
Intern-resident-attending identities	Y	Y	Y	Y	Y	Y
Time identities	N	Y	Y	N	Y	Y
Patient characteristics, service identities	N	N	Y	N	N	Y
Number of observations	151,765	151,765	151,765	151,765	151,765	151,765
Adjusted R^2	0.069	0.069	0.074	0.040	0.040	0.045
Sample outcome mean	4.027	4.027	4.027	123.28	123.28	123.28
Sample outcome standard deviation	1.342	1.342	1.342	177.83	177.83	177.83

Note: This table reports results for regressions on the effect of the community hospital rotation on intern spending, as in Equation (2.6.2). Models on the left panel are for log daily costs as the outcome variable; models on the right panel are for daily costs as the outcome variable. All models account for intern tenure and physician identities (intern-resident-attending interactions). Depending on the model, time identities (day of the week and month-year identities), patient characteristics (demographics, Elixhauser indices, Charlson comorbidity scores, and DRG weights), and admitting service identities (e.g., “cardiology team 1”) are also controlled for. Coefficients are listed with standard errors, clustered by admission, in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.7.1: Effect of Pre-training Characteristics and Other Hospital Training on Housestaff Spending

Characteristic of interest	Log daily test costs							
	Interns				Residents			
	(1) Male	(2) High USMLE	(3) Highly ranked	(4) Other hospital	(5) Male	(6) High USMLE	(7) Highly ranked	(8) Other hospital
Effect of characteristic	-0.021* (0.012)	-0.001 (0.012)	0.011 (0.018)	0.002 (0.025)	-0.040** (0.018)	-0.034* (0.020)	0.001 (0.031)	-0.169* (0.095)
Number of observations	186,186	190,777	131,366	220,074	176,939	190,777	110,898	220,074
Adjusted R^2	0.222	0.165	0.166	0.165	0.251	0.175	0.176	0.178
Sample outcome mean	4.028	4.028	4.043	4.022	4.028	4.028	4.043	4.022
Sample outcome standard deviation	1.341	1.341	1.335	1.353	1.341	1.341	1.335	1.353
Sample characteristic mean	0.596	0.277	0.226	0.049	0.560	0.286	0.205	0.060
Sample characteristic standard deviation	0.491	0.447	0.418	0.215	0.496	0.452	0.404	0.238

Note: This table reports results for regressions on the effect of the some pre-residency characteristics and of training at another academic hospital. All regressions are of the form in Equation (2.7.1), where the coefficient of interest is on a group of housestaff identified by either pre-residency characteristics or whether the housestaff is from the other academic hospital. The effect of many other characteristics of interest (or groups) were estimated and are described in Section 2.7.1; they are all insignificant and omitted from this table for brevity. Models on the left panel show results for groups of interns; models on the right show results for groups of residents. All models control for patient and admission characteristics, time dummies, and fixed effects for attending and the other housestaff on the team (e.g., the resident is controlled for if the group is specific to the intern). Standard errors are clustered by admission. * significant at 10%; ** significant at 5%; *** significant at 1%.

2.A Appendix: Tests for Exogenous Assignment

In this Section, I will present two sets of randomization tests for exogenous assignment. The evidence presented in here complements Tables 2.4.1 and 2.4.2 in the main chapter. Section 2.A.1 presents results regarding the assignment of patients to housestaff. Section 2.A.2 presents the assignment of other training experiences to housestaff: in particular, the assignment of supervising physicians and the community hospital rotation to housestaff.

2.A.1 Assignment of Patients to Housestaff

First, I test for the joint significance of housestaff identities in regressions of this form:

$$X_{ijkta} = \eta_t + \mu_a + \zeta_i^{(0)} + \zeta_j^{(1)} + \zeta_k + \varepsilon_{ijkta}, \quad (2.A.1)$$

where X_{ijkta} is some patient characteristic or linear combination of patient characteristics for the patient at a unique admission a at time t , being cared for by intern i , resident j , and attending k on the day of admission. η_t is the sum of time fixed effects, for the day of the week and the month-year interaction; μ_a is a fixed effect that corresponds to the admitting service (e.g., “heart failure service” or “oncology service”). $\zeta_i^{(0)}$, $\zeta_j^{(1)}$, and ζ_k are fixed effects for the intern i , resident j , and attending k , respectively. In the analyses that follow, I test for the joint significance of the fixed effects $(\zeta_i^{(0)}, \zeta_j^{(1)})_{i \in \mathcal{I}, j \in \mathcal{J}}$. For simplicity, I do not impose any relationship between the fixed effect of a housestaff as an intern and the fixed effect of the same housestaff as a resident.

In column (1) of Table 2.A.1, I show F -statistics and the corresponding p -values for the null hypothesis that $(\zeta_i^{(1)}, \zeta_j^{(2)})_{i \in \mathcal{I}, j \in \mathcal{J}} = \mathbf{0}$. I perform the regression (2.A.1) separately each of the following patient characteristics X_{ijkta} as a dependent variable: patient age, a dummy for male sex, and a dummy for white race. I also perform (2.A.1) using a linear prediction of log admission test spending, based on patient demographics, as the dependent variable. As shown in Table 2.A.1, I fail to find joint statistical significance for any of these tests.

Second, I test for the significance of housestaff characteristics in regressions of this form:

$$X_{ijkta} = \eta_t + \mu_a + \gamma_1 Z_i + \gamma_2 Z_j + \zeta_k + \varepsilon_{ijkta}. \quad (2.A.2)$$

Equation (2.A.2) is similar to Equation (2.A.1), except for the use of a vector of housestaff characteristics Z_i and Z_j for intern i and resident j , respectively. This specification tests more parsimoniously whether certain types of residents are more likely to be assigned certain types of patients. For housestaff h , I use a relatively rich set of characteristics Z_h : the housestaff’s position on the rank list; USMLE Step 1 score; sex; age at the start of training; and dummies for whether the housestaff graduated from a foreign medical school, whether he graduated from a rare medical school, whether he graduated from medical school as a member of the AOA honor society, whether he has a PhD or another graduate degree, and whether he is a racial minority.

Columns (2) and (3) of Table 2.A.1 show F -statistics and the corresponding p -values for the null hypothesis that $(\gamma_1, \gamma_2) = \mathbf{0}$. Column (2) includes all housestaff characteristics in Z_h ; column (3) excludes position on the rank list, since this information is missing for a sizeable proportion of housestaff. Patient characteristics for dependent variables in (2.A.2) are the same as in (2.A.1). Again, I fail to find joint significance for any of these tests.

Third, I plot the distribution of patient age and the predicted test costs across patients admitted to interns and residents with high or low estimated test spending practice styles. Figures 2.A.1 and 2.A.2 show kernel density plots of the age distributions for patients assigned to interns and residents, respectively, each of which compare housestaff with practice styles above and below the mean. Figures 2.A.3 and 2.A.4 are the corresponding figures plotting the distribution of predicted spending for patients assigned to housestaff with above- or below-mean spending practice styles. All of these figures show that there is essentially no difference across the distribution of age or predicted spending for patients assigned to housestaff with high or low spending practice styles. Kolmogorov-Smirnov statistics are also unable to reject the null that the underlying distributions are different.

2.A.2 Assignment of Other Training Experiences to Housestaff

In order to formalize inference for whether certain types housestaff are more likely to be assigned to certain types of housestaff and attending physicians, I perform the following regressions:

$$\hat{\zeta}_h^{(r)} = \gamma_h \hat{\zeta}_{-h}^{(1-r)} + \gamma_k \hat{\zeta}_k + \varepsilon_{ijka}, \quad (2.A.3)$$

where each observation corresponds to an admission a , but where error terms are clustered at the level of the intern-resident-attending team, since there are multiple observations for a given team. From a first stage, the estimated fixed effect for housestaff h is $\hat{\zeta}_h^{(r)}$, where $r = 0$ for interns and $r = 1$ for residents, and the estimated fixed effect for the other housestaff team member $-h$ is $\hat{\zeta}_{-h}^{(1-r)}$, where the superscript c denotes that the fixed effect for the same housestaff in different roles as intern and resident need not be the same. $\hat{\zeta}_k$ is the estimated fixed effect for attending k .²³ Estimates for γ_h and γ_k are small, insignificant, and even slightly negative.

Second, I perform a similar exercise as in the previous subsection, in which I plot the distribution of estimated attending fixed effects working with housestaff with above- or below-mean spending practice styles. Figures 2.A.5 and 2.A.6 show this for interns and residents, respectively. They show that the practice-style distribution for attendings is similar for those assigned to high- vs. low-spending housestaff. The Kolmogorov-Smirnov statistic rejects the null of equality when comparing the distributions for above-mean- and below-mean-spending residents; however, I am unaware of

²³I use two approaches to get around the reflection problem due to the first-stage joint estimation of $\zeta_i^{(0)}$, $\zeta_j^{(1)}$, and ζ_k (Manski, 1993). First, I perform (2.A.3) using “jack-knife” estimates of fixed effects, in which I exclude observations with $-h$ and k to compute the $\hat{\zeta}_h^{(r)}$ estimate that I use with $\hat{\zeta}_{-h}^{(1-r)}$ and $\hat{\zeta}_k$. Second, I use the approach by Mas and Moretti (2009), in which I include nuisance parameters in the first stage to absorb team fixed effects for (i, j, k) .

how to adjust the statistic for clustering at the team level, and therefore its statistical significance is overstated. More importantly, differences in the distributions do not appear quantitatively significant.

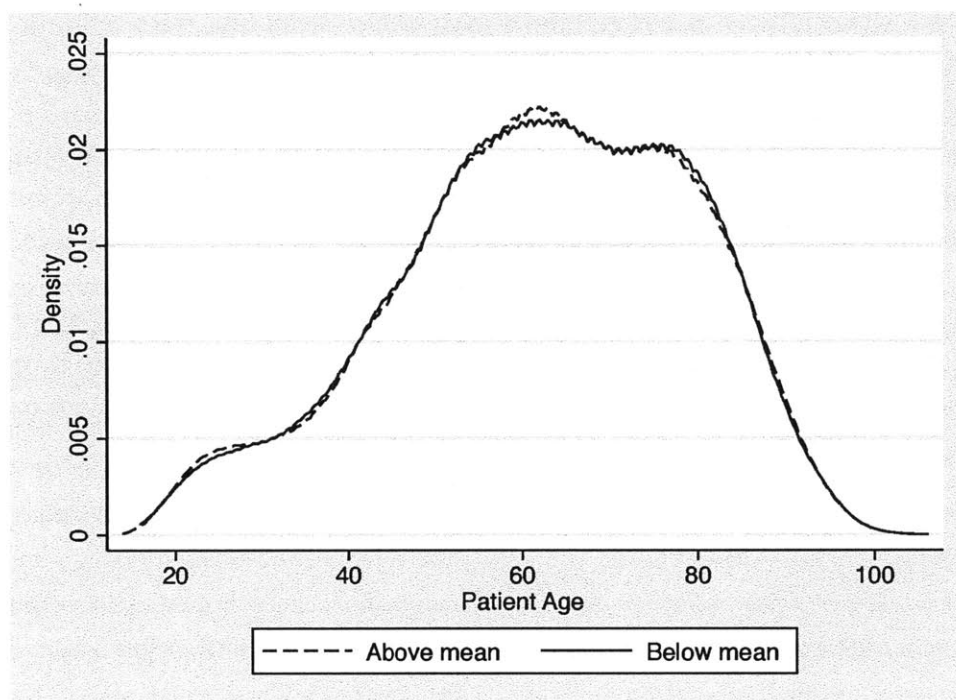
Finally, I formalize the inference that there is no significant difference in the scheduling of interns rotating to the community hospital. I focus on rotations to the community hospital because I show in the main chapter that this is the only rotation that significantly impacts housestaff practice styles. Recall that Table 2.4.1 in the main chapter shows that for interns who have not yet rotated to the community hospital in the first half of the year, the likelihood of rotating is similar for interns who developed a high practice style in the first half of the year and for those who developed a low practice style. I can formally show that the difference of 0.6% in the likelihood for a community-hospital rotation between these interns intern groups is statistically significant with a linear probability model (p -value = 0.925).

I also perform a new analysis on the time in which an intern first rotates to the community hospital. Denoting the first month of the community hospital for intern i as $FirstMonth_i$, I estimate this regression:

$$FirstMonth_i = \alpha + \gamma Z_i + \varepsilon_i,$$

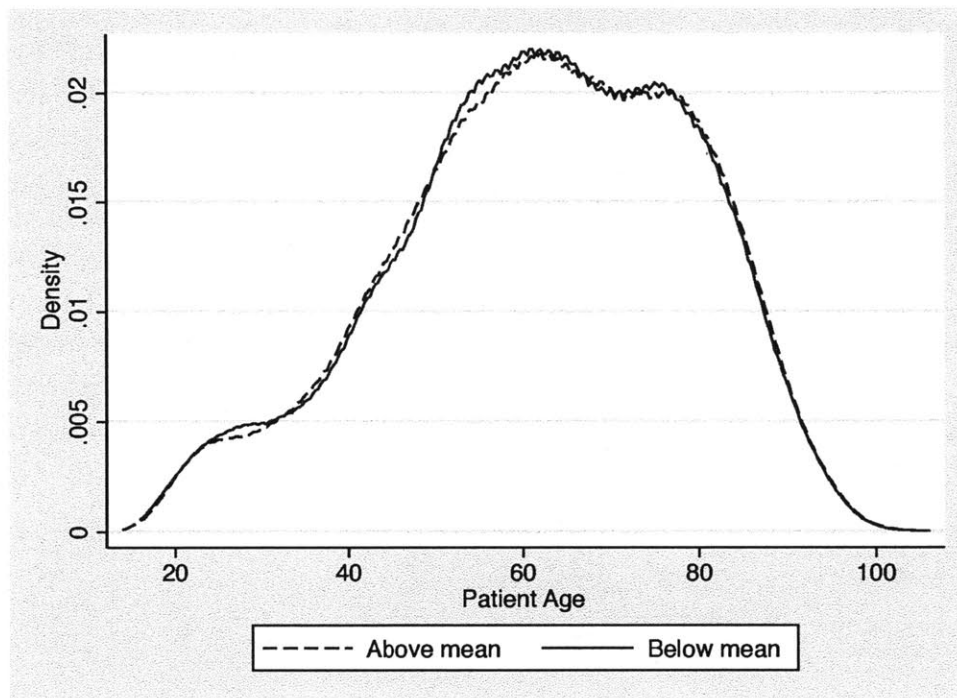
where Z_i is again a rich vector of characteristics for intern i , as described in the previous subsection. I test for the joint significance of $\gamma = 0$, and obtain an F -statistic of $F(10, 205)=0.95$, corresponding to a p -value of 0.492. Thus, I cannot reject the null that certain types of interns are more likely to rotate to the community hospital earlier.

Figure 2.A.1: Age of Patients Assigned to Interns with Above-mean vs. Below-mean Spending Practice Styles



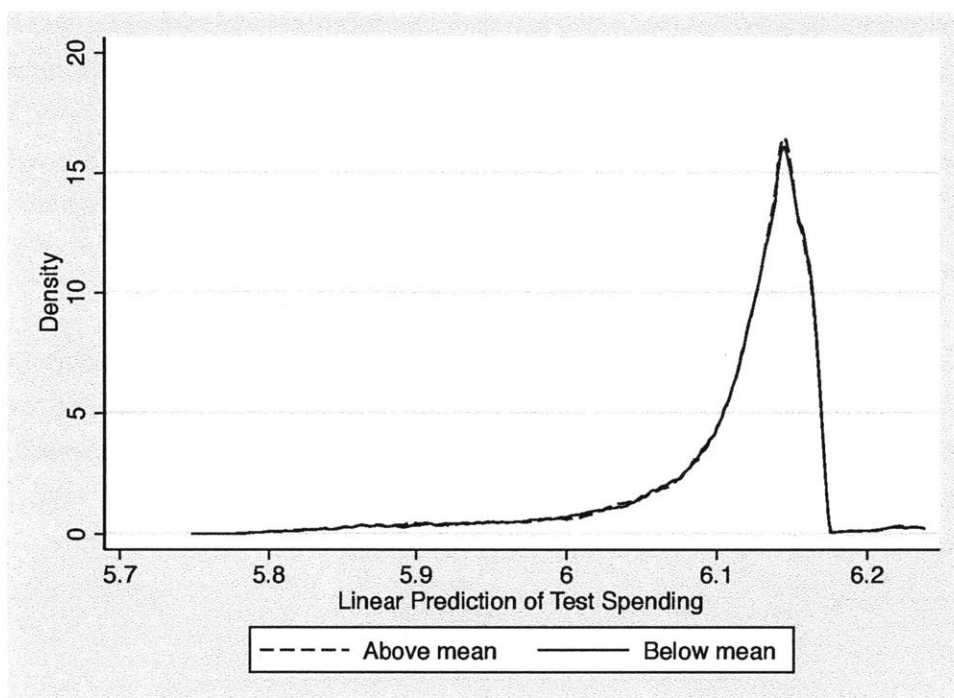
Note: This figure shows a kernel density plot of the distribution of patient age for patients assigned to interns (first-year housestaff) with test spending practice styles above the mean and those with practice styles below the mean. The density for above-mean interns is shown in dashed blue; the density for below-mean interns is shown in solid red. Physician productivity is estimated by fixed effects in a regression of log test spending, controlling for patient demographics (age, sex, and race) and comorbidities (Elixhauser indices and Charlson scores); time dummies (month-year combination and day of the week for the admission).

Figure 2.A.2: Age of Patients Assigned to Residents with Above-mean vs. Below-mean Spending Practice Styles



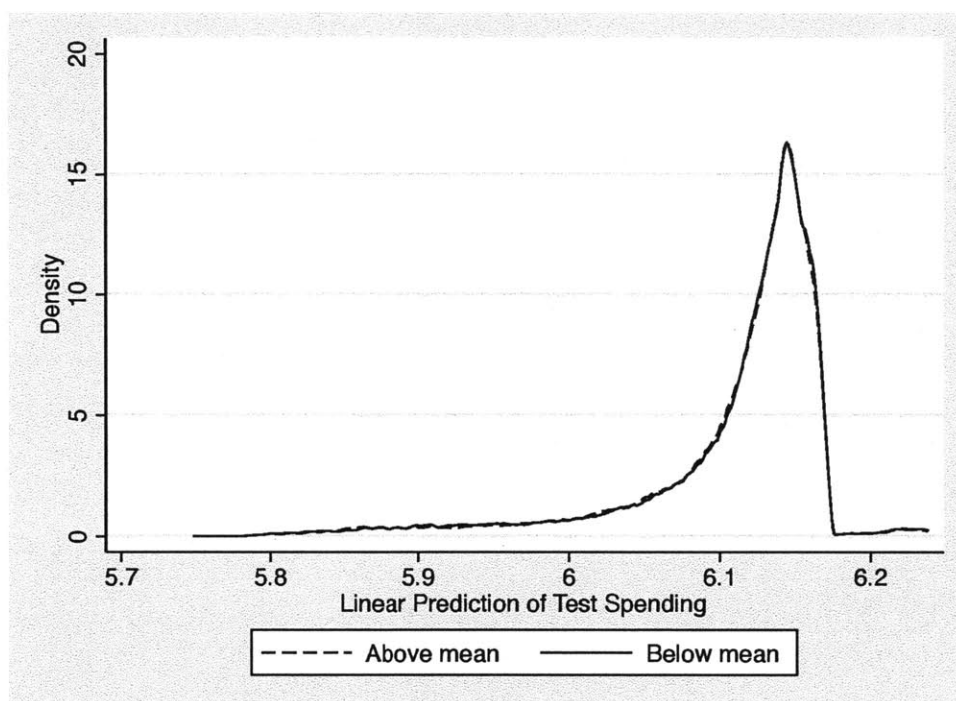
Note: This figure shows a kernel density plot of the distribution of patient age for patients assigned to residents (second- or third-year housestaff) with test spending practice styles above the mean and those with practice styles below the mean. The density for above-mean residents is shown in dashed blue; the density for below-mean residents is shown in solid red. Physician productivity is estimated by fixed effects in a regression of log test spending, controlling for patient demographics (age, sex, and race) and comorbidities (Elixhauser indices and Charlson scores); time dummies (month-year combination and day of the week for the admission).

Figure 2.A.3: Predicted Spending for Patients Assigned to Interns with Above-mean vs. Below-mean Spending Practice Styles



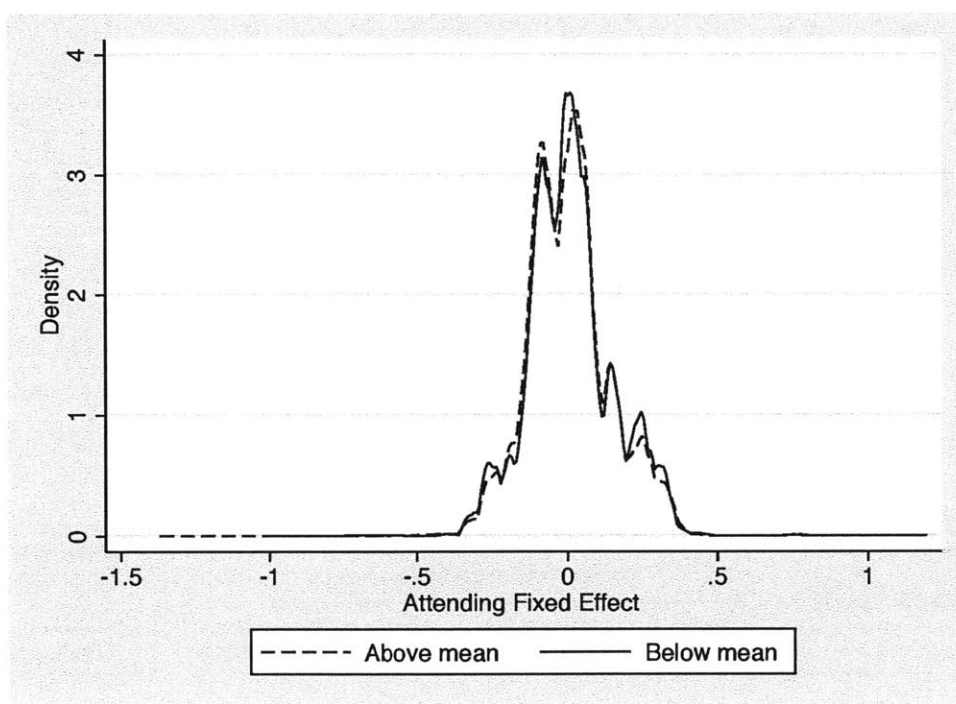
Note: This figure shows a kernel density plot of the distribution of predicted test spending for patients assigned to interns (first-year housestaff) with test spending practice styles above the mean and those with practice styles below the mean. The density for above-mean interns is shown in dashed blue; the density for below-mean interns is shown in solid red. Predicted spending is based on a simple linear prediction of total admission spending from patient age, age squared, sex, and race. Physician productivity is estimated by fixed effects in a regression of log test spending, controlling for patient demographics (age, sex, and race) and comorbidities (Elixhauser indices and Charlson scores); time dummies (month-year combination and day of the week for the admission).

Figure 2.A.4: Predicted Spending for Patients Assigned to Residents with Above-mean vs. Below-mean Spending Practice Styles



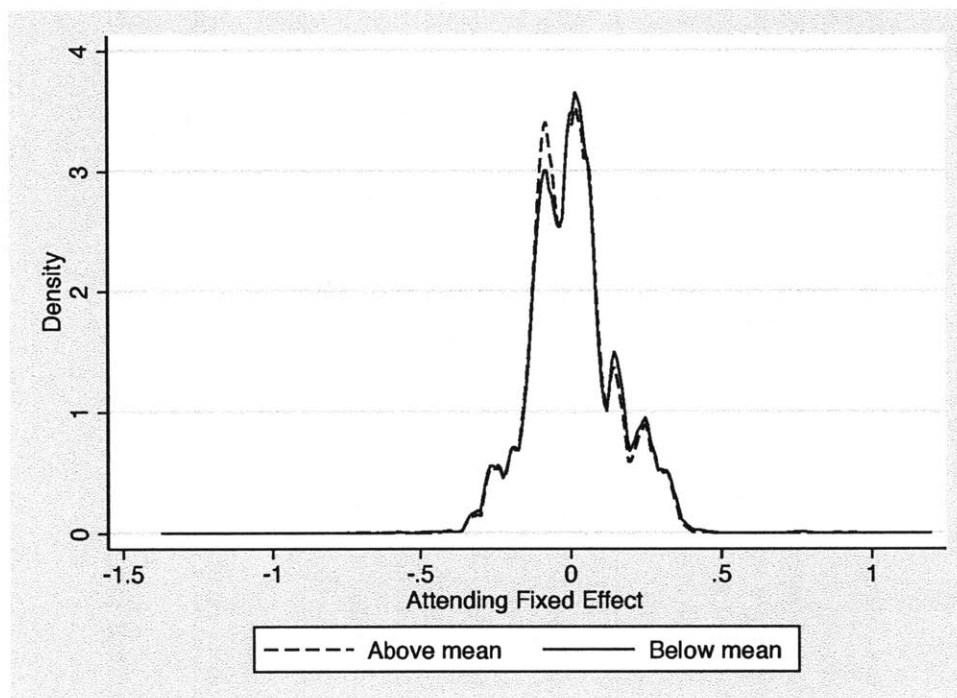
Note: This figure shows a kernel density plot of the distribution of predicted test spending for patients assigned to residents (second- or third-year housestaff) with test spending practice styles above the mean and those with practice styles below the mean. The density for above-mean residents is shown in dashed blue; the density for below-mean residents is shown in solid red. Predicted spending is based on a simple linear prediction of total admission spending from patient age, age squared, sex, and race. Physician productivity is estimated by fixed effects in a regression of log test spending, controlling for patient demographics (age, sex, and race) and comorbidities (Elixhauser indices and Charlson scores); time dummies (month-year combination and day of the week for the admission).

Figure 2.A.5: Spending Practice Styles for Attendings Assigned to Interns with Above-mean vs. Below-mean Spending Practice Styles



Note: This figure shows a kernel density plot of the distribution of spending practice styles for attendings assigned to interns (first-year housestaff) with test spending practice styles above the mean and those with practice styles below the mean. The density for above-mean interns is shown in dashed blue; the density for below-mean interns is shown in solid red. Physician productivity is estimated by fixed effects in a regression of log test spending, controlling for patient demographics (age, sex, and race) and comorbidities (Elixhauser indices and Charlson scores); time dummies (month-year combination and day of the week for the admission).

Figure 2.A.6: Spending Practice Styles for Attendings Assigned to Residents with Above-mean and Below-mean Spending Practice Styles



Note: This figure shows a kernel density plot of the distribution of of spending practice styles for attendings assigned to residents (second- or third-year housestaff) with test spending practice styles above the mean and those with practice styles below the mean. The density for above-mean residents is shown in dashed blue; the density for below-mean residents is shown in solid red. Physician productivity is estimated by fixed effects in a regression of log test spending, controlling for patient demographics (age, sex, and race) and comorbidities (Elixhauser indices and Charlson scores); time dummies (month-year combination and day of the week for the admission).

Table 2.A.1: Tests of Joint Significance of Housestaff Identities and Characteristics

Patient characteristic	Independent variables		
	Housestaff identities (1)	Housestaff characteristics (2)	(3)
Age	$F(1055, 46364) = 0.98$ $p = 0.655$	$F(20, 16069) = 0.68$ $p = 0.848$	$F(18, 37494) = 0.79$ $p = 0.711$
Male	$F(1055, 46364) = 1.01$ $p = 0.389$	$F(20, 16069) = 1.18$ $p = 0.256$	$F(18, 37494) = 1.26$ $p = 0.201$
White	$F(1055, 46364) = 1.02$ $p = 0.356$	$F(20, 16069) = 0.79$ $p = 0.734$	$F(18, 37494) = 0.92$ $p = 0.558$
Predicted spending	$F(1055, 46364) = 0.98$ $p = 0.685$	$F(20, 16069) = 0.79$ $p = 0.731$	$F(18, 37494) = 1.08$ $p = 0.368$

Note: This table reports tests of joint significance corresponding to Equations (2.A.1) and (2.A.2). Column (1) corresponds to Equation (2.A.1); columns (2) and (3) correspond to (2.A.2). Column (2) includes all housestaff characteristics: housestaff's position on the rank list; USMLE Step 1 score; sex; age at the start of training; and dummies for whether the housestaff graduated from a foreign medical school, whether he graduated from a rare medical school, whether he graduated from medical school as a member of the AOA honor society, whether he has a PhD or another graduate degree, and whether he is a racial minority. Column (3) includes all housestaff characteristics except for position on the rank list. Rows correspond to different patient characteristics as the dependent variable of the regression equation; the last row is predicted test spending using patient demographics (age, sex, and race). F -statistics and p -values are reported for each joint test.

Chapter 3

The Impact of Charging Low-income Families on Health Insurance Choices

As health care reform is implemented in the United States, coverage for low-income populations will be expanded not through a traditional public insurance model, but rather through an “defined contribution exchange” mechanism. Under this approach, low-income individuals would have a choice of a number of options for their insurance coverage. Individuals would receive a subsidy to purchase insurance that was tied to the lowest-cost plan (or some index of low-cost plans) and would pay some part of the difference if they chose a more expensive plan.

This major departure from the traditional free/single-choice public payer model raises a number of important questions. Will low-income consumers benefit from a wide choice of plans, or will “choice overload” or other problems discussed in the growing behavioral economics literature lead to worse outcomes than with a single option? Would charging the full differential between plans lead to adverse selection problems of the type discussed by Cutler and Reber (1998) in their study of Harvard University’s health insurance plan, or would those problems be minimized in this highly subsidized environment? Will shopping across plans lead to cost-reducing supply side competition among insurance companies or just increased efforts to find the lowest-risk cases?

Answering many of these questions requires addressing a key initial question: How price-sensitive will low-income consumers be in choosing across plans? If price sensitivity is low, for example, then choice is less likely to lead to cost-reducing competition among plans. If price sensitivity is very differential by enrollee health, then it suggests that plans that have different prices may also enroll patients of different levels of health. Specifically, if healthy patients are very price-sensitive, then we may have cheap plans attract healthy patients, with sicker patients remaining in more expensive plans. If adverse selection does play such a role in the insurance market, then more expensive plans with sicker enrollee populations would be unable to compete with cheaper plans with healthier populations.

In this paper, we study the plan choice of low-income enrollees in Massachusetts’ Commonwealth

Care program that was established as part of the state's health reform in April, 2006. Enrollees in Commonwealth Care were given a choice of up to four Medicaid Managed Care Organizations (MMCOs) from which they could receive their coverage. For about half of enrollees (those below the poverty line), this decision had no financial implications. But for the remainder, enrollees were charged not only a base contribution rate, but the differential cost of their plan choice over the lowest-cost plan in their area. The financial implications of this decision were non-trivial; the average differential in 2007 between the non-lowest-cost plan and the lowest-cost plan across areas and income groups was \$18.83 per month, and the maximum was \$116 per month. For the relevant population, those with incomes from 100% to 300% of poverty line, these are meaningful amounts.

Most relevantly for our study, there was a major shift in the pricing of plans for open enrollment in June 2008. Before that time, all enrollees below 150% of the federal poverty line received health insurance for free regardless of cost. However, after June 2008, enrollees between 100% and 150% of poverty were responsible for the full differential in cost between the lowest-cost plan and the plan that they choose. In addition to this major change, most other plan types for individuals above 150% of poverty also experienced an increase in the base contribution rate that enrollees were responsible for. The shifts in base contributions occurred by state mandate, independent of enrollee and MMCO actions.

We have created a unique data set using information from the state of Massachusetts on the enrollment decision of each Commonwealth Care recipient over the 2007-2008 period. We have data both on those enrollees who were already in a plan as of June 2007 and faced the decision over whether to switch plans ("old enrollees"), and those who were newly choosing across plans through 2008 ("new enrollees"). For each enrollee we have information on their income and an index of their underlying medicalspending risk. We estimate a conditional logit model on these data, using the shift in pricing described above, to understand how price differentials impact plan choice.

We find that these low-income populations are highly sensitive to plan price differentials. Our central estimate suggests that a \$10 increase in out-of-pocket costs for a given plan in 2008 decreases enrollment by 10.1% for old enrollees who are deciding whether to switch plans and by 17.2% for new enrollees who are deciding on a plan after entering Commonwealth Care. These are indeed substantial numbers of enrollees for a relatively small difference in price. Because out-of-pocket costs are low in general and even free for many consumers, the implied price elasticities of demand are -0.647 for old enrollees and -0.714 for new enrollees.

Our paper proceeds as follows. Section 3.1 describes the institutional background on Commonwealth Care and the price changes we study. Section 3.2 discusses our data and empirical approach. Section 3.3 presents results, and Section 3.4 concludes.

3.1 Institutional Background of Commonwealth Care

The groundbreaking health care reform passed in Massachusetts in 2006 had a number of important features, including a mandate on individuals to purchase insurance and a reform of non-group and small group insurance markets. Most important for our purposes, the law established the Commonwealth Care program for those in families with incomes below three times the poverty line (roughly \$30,000 for singles and \$60,000 for a family of four at the time of the law's passage). Only individuals who were not eligible for other coverage (employer-sponsored insurance or Medicaid) could enroll.

Starting in mid-2007, the first full year of the program, individuals were placed in one of six “plan types” depending on their income. Plan types were differentiated by the patient cost-sharing imposed in the plan and by enrollee contribution rates. Those below poverty were in plan type I; those who were 100-150% of poverty were in plan type IIA; and those who were 150-200% of poverty were in plan type IIB. For those 200-300% of poverty, there was in 2007 a choice of two different benefits structures, with plan type III having higher copayments and a lower premium cost to enrollees, and plan type IV having lower copayments and a higher premium cost to enrollees; within plan types III and IV there was a division into IIIA/B and IVA/B at 250% of poverty.

Enrolling individuals had a choice of up to four Medicaid Managed Care Organizations (MMCOs); in some areas of the state the choice set was smaller due to limited regional coverage of some MMCOs. For example, by 2008, enrollees in Western Massachusetts generally only had 2 choices, while enrollees in Northern Massachusetts predominantly had 4 choices and enrollees elsewhere, including Boston, had on average 3 choices.

In 2007, individuals below 150% of the poverty line were enrolled in plan type I or plan type IIA for free, and could enroll in any of the available MMCO at no personal cost. Individuals in the remaining plan types had to pay a base contribution for the lowest-cost plan available in their area, as well as paying the full differential in the cost of choosing any plan that was above that lowest-cost plan. The base contribution was \$35 for plan type IIB, \$70 for PT IIIA, and \$105 for PT IIIB. Enrollees choosing plan types IVA and IVB had the same base contributions as those in plan types IIIA and IIIB, respectively, since they were of the same income groups, but they because they chose the lower copayment plans, out-of-pocket contributions for all plans in IVA or IVB were above the base contribution. Other than plan premiums, copayments within each plan type were standardized across MMCOs by the state.

In addition to enrollees who explicitly chose a plan, individuals below 150% of the poverty line who were deemed eligible but did not choose a plan were auto-enrolled. The auto-enrollment algorithm could randomize auto-enrollees among several possible plans but was weighted towards low-cost plans. The MMCOs themselves made bids in early 2007 for the prices they would charge the state for each demographic group. Costs were calculated for each region based on the demographic composition of each region. Incentives to bid low came from both the assignment of auto-enrollees

and the financial incentives for enrollees to choose low-cost plans.

The system then changed in several important ways for open-enrollment in June, 2008. First, those in plan type IIA still could sign up for the lowest-cost option for free, but now had to pay the full differential for choosing a more expensive plan in their area. Second, plan types IVA and IVB, the more expensive plans with lower copayments for individuals at 200-300% of FPL, were discontinued. Once again, MMCOs made bids for each plan type and demographic group understanding these structural changes in choice incentives.

The result was a very significant shift in the cost of plan enrollment for those with incomes above the poverty line. These changes are illustrated in Table 3.1.1. This Table shows, for each plan type, the mean and standard deviation of the change in contribution for the typical enrollee to stay in the same plan in 2008 as in 2007, as well as the mean and standard deviation of the change in contribution for the typical enrollee to move from the lowest- to highest-priced plan in the area. Enrollees in plan types IIA and IIB experienced an average increase in contributions of \$9.14 and \$14.04, respectively, while enrollees in plan types IIIA and IIIB experienced an increase in contributions of \$48.66 and \$50.85, respectively. While plan types IVA and IVB were discontinued in 2008, we present 2008 figures for increases in out-of-pocket costs assuming that enrollees would continue in the same MMCO in plan types IIIA and IIIB, respectively. Average contribution increases are \$16.08 and \$21.47, for enrollees originally in plan types IVA and IVB, because these plans were generally more expensive than plan types IIIA and IIIB, respectively, in 2007. In the next set of columns in the Table, we show that the difference between the highest- and lowest-priced choices that an enrollee faced generally increased from 2007 to 2008. Most notable are enrollees in plan type IIA, who faced no price differential in 2007 and then a \$24.15 differential in 2008.

One important task to assess the welfare implications of this program is to understand the mechanism by which the MMCOs chose their income-group and area-specific contribution prices. In this study, we take those prices as given and exogenous to the decisions made by enrollees. Future work could usefully explore the price setting side of this policy change as well.

3.2 Data and Empirical Strategy

In order to assess the impact of this change in relative plan prices on plan choice, we have collected two sets of data with the helpful assistance of the staff of the Massachusetts Health Connector. The first set of data is information on all those who were enrolled in Commonwealth Care continuously from June 2007 to September 2008, a total of 75,184 “old enrollees.” Table 3.2.1 shows the number of enrollees in each plan type and MMCO for both 2007 (as of June 2008 prior to open enrollment) and in September 2008 (after open enrollment), by plan type and MMCO. Among the MMCOs, we see that Boston Medical Center HealthNet Plan (BMC) had the largest group of enrollees, while Fallon Community Health Plan (Fallon) was just entering the market at this time. Also, the greatest proportion of enrollees above poverty was just above poverty, in plan type IIA or at 100-150% of

poverty.

For each old enrollee, we have data on demographic characteristics (age and sex), health care utilization, area of residence, original plan choice in 2007, and new plan choice in 2008. Area of residence was categorized by the Commonwealth Connector Authority into 5 regions and 38 areas. Of note, we have three actuarial measures of health risk for old enrollees, one based on enrollee demographics, another based on health care utilization, and a third based on a combination of enrollee demographics and health care utilization.

The second dataset contains information on all first-time enrollees during 2008, which includes a total of 115,010 “new enrollees,” of whom 52,307 are above poverty and therefore faced financial consequences in choosing a health plan. Table 3.2.2 presents similar information on new enrollment in 2008 as does Table 3.2.1, breaking down numbers of new enrollees by plan type and MMCO. BMC again comprises the largest group of new enrollees, and Fallon has the largest increase in enrollees from 2007 to 2008. Table 3.2.3 presents new enrollment by month and plan type. We find that new enrollment decreased over time, which is not surprising given that the number of uninsured in Massachusetts also decreased over time. For each new enrollee, we have similar data as for old enrollee, with the exception of health utilization data, since they are enrolling in Commonwealth Care for the first time.

Finally, our third set of data is at the plan level categorized by area, plan type, and MMCO. Compiling out-of-pocket contributions for each plan, for each enrollee we constructed a choice set with corresponding contribution prices for each MMCO in the choice set, depending on the enrollee’s income level and area of residence. Using this information, we constructed conditional logit models to describe the discrete choice that old and new enrollees faced. For old enrollees, we modeled the utility of a plan choice in 2008 for an enrollee as a function of whether the enrollee had chosen that plan previously in 2007, the contribution price of the choice in 2008 during open enrollment, the previous contribution price of the choice in 2007, enrollee health risk and income level (plan type), and MMCO dummies:

$$u_{ij}^{old} = \beta_1 Same_{ij} + \beta_2 P_{08,j} + \beta_3 P_{07,j} + \beta_4 P_{08,j} Health_i + \beta_5 P_{08,j} PT_i + \beta_6 Plan_j + \varepsilon_{ij},$$

where i indicates the enrollee and j indicates the plan. $Same_{ij}$ is a dummy for whether old enrollee i was enrolled in plan j in 2007; $P_{08,j}$ and $P_{07,j}$ are contribution prices for plan j in 2008 and 2007, respectively; $Health_i$ and PT_i are vectors of dummies corresponding to quartiles of health risk and plan type for enrollee i , respectively; $Plan_j$ is a vector of dummies for each MMCO; and the error term ε_{ij} is identically and independently distributed as extreme value. Of note, we can use more than one measure of health risk for old enrollees, as described above, since we have information on both demographics and health care utilization. Also, although the enrollee’s 2007 plan type can be thought of as equivalent to the enrollee’s income bracket, for enrollees at 200-300% of poverty, it also communicates additional information on enrollee preferences in 2007 between a high-copayment,

lower-premium arrangement (plan types IIIA and IIIB) and a low-copayment, higher-premium one (plan types IVA and IVB). When relevant, we normalize the health-risk quartile dummies with respect to the first quartile, plan-type dummies with respect to plan type IIA, and plan dummies with respect to BMC.

For new enrollees, we use a similar conditional logit model, although one that is slightly simpler than for old enrollees because we have less information:

$$u_{ij}^{new} = \gamma_1 P_{08,j} + \gamma_2 P_{08,j} + Health_i + \gamma_3 P_{08,j} PT_i + \gamma_4 Plan_j + \varepsilon_{ij},$$

where variables are similar to those defined above for old enrollees. Because new enrollees are enrolling for the first time, information on previous plan or the contribution price in 2007 is not relevant. Relatedly, plan type is now equivalent to income level, because we do not have information whether a new enrollee at 200-300% of poverty would have preferred a low-copayment or high-copayment plan type, as these were discontinued in 2008. Finally, because we do not have information on health care utilization in the previous year, we have only one measure of health risk, based on demographics.

Based on the above conditional logit models, we estimate the likelihood for each enrollee facing a given choice set to enroll in a given plan. We then simulate pricing counterfactuals in which a plan's contribution price is increased by \$10, and we also calculate corresponding elasticities for a percentage increase in contribution price. We compare price responses and elasticities both for old enrollees and for new enrollees and further subgroup responses by health risk and by income level. For old enrollees, we also examine counterfactual responses for various types of old enrollees independent of health risk and income level, such as those who initially chose the lowest-price plan or those who chose the highest-price plan in 2007, or those who ex post stayed in or switch their previous plans.

3.3 Results

3.3.1 Conditional Logit Regressions

We first present results of conditional logit regressions for old enrollees and for new enrollees. In addition to full regressions as described above, we also estimated restricted regressions in which we only focus on certain variables, such as 2008 contribution price, while leaving out health-risk and income interaction terms. Table 3.3.1 summarizes our regression results.

The first four columns in Table 3.3.1 present coefficients for regressions of old enrollee choices. The first column represents the simplest model, which focuses on the same-plan dummy and contribution price in 2008 but also includes plan dummies. In this restricted model, we find that whether a plan was chosen in the previous year has by far the most significant effect on which plan the old enrollee chooses in 2008. Other coefficients are also significant, including the negative coefficient

on price. But the relative magnitudes of the same-plan dummy coefficient and the price coefficient imply that a price difference of about \$82 would be required for an old enrollee to be equally likely to switch out of his old plan, all else equal.

The next column in Table 3.3.1 presents coefficient results when we add the contribution price from 2007, when the old enrollee last chose a plan. We find that a significant positive coefficient on this previous-year price that has a magnitude that is about 60% of that of the negative coefficient on current-year price. A couple observations can be made with respect to the coefficient on previous-year price. First, the fact that it is not zero suggests that price may be positively correlated with unobservable plan characteristics that may increase enrollee utility. Because copayments are the same across plans in each enrollee's choice set, these characteristics may include the number of physicians or hospitals accepting a given plan. Second, current-year price is greater in magnitude than previous-year price, which implies that price still matters despite unobservable characteristics. That is, in equilibrium, lower-priced plans will have greater observed demand, even if prices do not change from 2007 to 2008.

The third and fourth columns in Table 3.3.1 present coefficients for the full conditional logit model, which include interactions between price and health and between price and plan type. Health is considered in quartiles, with the highest quartile representing the healthiest (lowest-risk) patients and the first quartile representing the sickest (highest-risk) patients. The third column uses demographic health risk and does not find a significant or monotonic relationship between price sensitivity and health. On the other hand, when we consider health risk based on utilization data, price sensitivity does have a significant and monotonically decreasing relationship with health (recall that the coefficient on price is negative, so that increasingly positive coefficients on interaction terms imply a decreasing effect of price). Although not presented in Table 3.3.1, we also fit the model using the comprehensive measure of risk based on both demographics and utilization and do not find any improvement in explanatory power or a change in the coefficients. We conclude that most if not essentially all the useful information on old-enrollee choices according to health is captured by utilization data and not demographics.

In both columns 3 and 4, we find a significant change in price sensitivity according to income level or plan types. Enrollees in higher income groups are less price sensitive than those in lower income groups. Also, within income groups at 200-300% of poverty, plan-type choice in 2007 between low-premium (plan types IIIA and IIIB) and high-premium (plan types IVA and IVB) also predicts price sensitivity in 2008 when the choice is no longer available: enrollees who chose the high-premium, low-copayment plans are much less price-sensitive than enrollees with similar incomes but with the opposite choices in the previous year. Also, in both columns, the effect of income on price sensitivity is much greater than the effect of health on price sensitivity.

Finally, the last two columns in Table 3.3.1 present regression results for new enrollees. As mentioned above, the conditional logit regressions using new-enrollee data are necessarily simpler, since there are no previous-year choices. Coefficients cannot be meaningfully compared with results

for old enrollees, because the scale is necessarily different without the presence of a large coefficient on the same-plan dummy. However, we can compare the relative magnitudes of coefficients within columns: Specifically, we note that, in the last column for new enrollees, coefficients on price-health interactions are larger relative to the price coefficient or the price-income interactions, as compared to either the third or fourth columns for old enrollees. Moreover, the health measure in the new-enrollee regression is demographic-based, since we do not have utilization data on new enrollees. Therefore, we can conclude that health influences price sensitivity to a larger extent for new enrollees, who have not been insured and perhaps have not been connected to the health care system, than for old enrollees. On the other hand, the effect of income on price sensitivity is relatively small and statistically insignificant.

3.3.2 Predicted Responses to Price Changes

The conditional logit regressions above describe enrollee choice as likelihoods that depend on the prices of options in the choice set, enrollee health and income, and in the case of old enrollees previous choices. We will now use these models to predict demand responses to price changes. There are several ways to conceptualize responses to price changes. First, because prices for all options in the choice set determine choices, responses can be conceptualized as own-price and cross-price responses in permutations between each one of the four MMCOs. Second, because many of the plans are free to enrollees, traditional measures of price elasticities that consider percent changes in price may fail to capture responses to price; however, prices are unlikely to increase as much in the low plan types as they would in the higher plan types. Therefore, we may consider both traditional elasticities and responses to absolute increases in price. Third, we are interested in the relative price responses among different subgroups of enrollees, for example for healthy or sick patients or for low-income or high-income patients. We could theoretically present all permutations of own-price and cross-price responses for each of these subgroups, but for simplicity we will present only own-price elasticities averaged over the four MMCOs by their respective proportional enrollments.

In Table 3.3.2, we present own-price and cross-price responses in enrollment to a \$10 increase in the price of each of the four MMCOs, for old enrollees in the top panel and for new enrollees in the bottom panel. In each panel, columns represent the MMCO that undergoes the \$10 price increase, and rows represent the MMCO that is impacted by the price change; own-price responses are displayed on the diagonal, and cross-price responses are off the diagonal. Responses are aggregated across plan types and areas. Cross-price responses are stronger when plans with greater initial enrollment, BMC and Network Health, undergo a price change, because proportional movements of enrollees from these plans would be greater in magnitude. At the same own-price responses are relatively low for these larger plans, while higher for the smaller plans of Fallon and Neighborhood Health Plan (NHP). For old enrollees, this is due to the fact that since these plans are larger, a larger proportion of potential enrollment comes from those already previously enrolled, who are much more likely to stay than to switch despite price changes. Therefore, we see that own-price

responses for these plans are relatively higher for new enrollees (although still lower than for the other two smaller plans, due to higher plan dummies). Finally, we note that responses are much greater in magnitude for new enrollees than for old enrollees, which again results from the fact that previous enrollment dampens switching responses through the propensity to stay in the same plan despite unfavorable price differentials.

In Table 3.3.3, we present own-price and cross-price elasticities of demand, for old enrollees in the upper panel and for new enrollees in the lower panel. As we do for price responses to an absolute change in price, we average elasticities over enrollees across different plan types and areas. We find similar patterns as in Table 3.3.2. However, some responses seem larger (or smaller) when presented as elasticities, for example the response in Fallon enrollment to a change in Network Health's price. These differences arise from the fact that some plans are priced relatively higher than others, which means that a percentage increase in price translates to greater or lesser absolute price changes. Similarly, we note that unlike responses to the absolute price change in Table 3.3.2, elasticities for new enrollees are not as significantly larger in magnitude. This can be understood by the following reasoning: Elasticities are greater in magnitude when prices are higher, but plans with higher prices are chosen less frequently, so that weighted averages of elasticities will have smaller differences or even differences of the opposite sign when compared to weighted averages of absolute price responses.

Of note, the conditional logit model that presumes independence of irrelevant alternatives (IIA) and therefore constructs identical cross price elasticities for any alternative with respect to a given good changing prices. As such, an elasticity table for a single enrollee should have identical cross-price elasticities within each column. However, we do not see average cross-price elasticities identical within columns because, although cross-price elasticities are identical for each enrollee, enrollees have different characteristics (for example previous plan choice) and sometimes different choice sets, such that averages across enrollees are not necessarily identical. Although we do not present the results, we also estimated price elasticities by using the restricted models of column 1 and 5 in Table 3.3.1 and do not find that average elasticities change by much. We therefore conclude that most of the variance in average cross-price elasticities results not from heterogeneity in enrollees health or income across MMCOs but from plan fixed effects, previous plan choice, and different choice sets for each area and plan type.

Finally, we consider price responses among different subgroups of enrollees in Tables 3.3.4 and 3.3.5. In comparing subgroups we focus on aggregate own-price responses and ignore cross-price responses that we have considered above. Table 3.3.4 presents results for a \$10 absolute increase in price, and Table 3.3.5 presents own-price elasticities. In both tables, we present results for various populations in columns, and we subgroup each population in rows according to health and plan type.

We find variation in price responses to a \$10 price increase by health and by plan type both for the overall population of old enrollees (columns 1 and 2 in Table 3.3.4) and for new enrollees

(column 7 in Table 3.3.4). For the overall population of old enrollees, we consider both demographic- and utilization-based measures of health, and as we have reported for the regression coefficients, utilization-based measures perform better at capturing the variation, although for new enrollees the demographic-based measure of health seems to predict variation better than either measure does for the old enrollees. On the other hand, variation is more discernible by plan type among old enrollees than among new enrollees. Moreover, preference for low-copayment, higher-premium plan types in the previous year is a much stronger predictor of price responsiveness than any income grouping, even though such an option was no longer in enrollee choice sets during the period that we study; enrollees with such preferences (those who previously chose plan types IVA or IVB over plan types IIIA or IIIB) exhibit roughly half the price responses than enrollees who preferred otherwise. When considering elasticities for the same enrollees (the same columns in Table 3.3.5), however, these relationships are obscured or even seemingly reversed. For example, elasticity patterns according to health are less clear, and enrollees are actually more price elastic with higher income groups, which is due to more expensive plans for higher income groups.

Since we have information on previous choices for old enrollees, we were also able to examine price responses for different populations of old enrollees according to their previously revealed preferences. We first considered old enrollees who chose the cheapest or most expensive plans in 2007 (columns 3 and 4, respectively, in both Tables 3.3.4 and 3.3.5). Old enrollees who chose the lowest-priced plan in their 2007 choice set are more price sensitive than those who chose the highest-priced plan in 2007, by both absolute-price increase and elasticity metrics and in each subgroup. Especially impressive differences are revealed when comparing these enrollees within subgroups who chose out of plan types III or IV. We then considered old enrollees according to whether they stayed in the same plan or switched plans from 2007 to 2008. At first glance, this selection of populations seems endogenous to the question of price responsiveness that we want to ask; however, it is important to note that this selection is based on ex post decisions, while predicted price responses are based on counterfactual price differences. Selecting on these populations, we find that enrollees who stayed in their previous plan ex post have lower responses than those who switched plans ex post, both in terms of absolute-price responses and elasticities.

3.4 Conclusion

Using a unique dataset that includes all enrollees in the Massachusetts Commonwealth Care program during 2008, we examine how the availability of healthplan choice and financial responsibility affect the decisions of low-income families seeking subsidized health insurance. Specifically, we ask by how much low-income consumers respond to differences in price, and whether price sensitivity differs among them. We find that low-income consumers are indeed price-sensitive. A \$10 increase in the monthly premium of a plan will on average decrease enrollment by 10.1% for old enrollees potentially switching plans and by 17.2% for new enrollees who have decided to enroll for the first

time. Own-price elasticities of -0.647 and -0.714 for old enrollees and new enrollees, respectively, belie the fact that many of the plans that we study are quite cheap and some are even free, so that a percentage increase in price for these plans may represent small increases. Indeed at higher prices, such as those for enrollees in higher income groups, elasticities are significant.

Behind these average numbers, there is also evidence of substantial heterogeneity in price sensitivity. Some of this heterogeneity can be explained by health-spending risk, and some is also associated with income, but we find that there is likely even greater heterogeneity that is revealed by previous choices (such as whether the enrollee preferred a high-copayment or low-copayment plan, or whether the enrollee preferred the cheapest plan in his choice set or the most expensive one) which are otherwise not obvious from enrollee demographics or income. Moreover, old enrollees who are deciding whether to stay in their current plan or switch to another one behave very differently from new enrollees who have never been enrolled before. Old enrollees are extremely likely to stay in their current plan, requiring a monthly price difference between that plan and an alternative of about \$82 to be indifferent between staying in the current plan and switching to the alternative. Consequently, they are much less price-sensitive than new enrollees. In particular, new enrollees seem to respond more differently to prices according to health status, with healthier enrollees being more price sensitive, than do old enrollees.

These results suggest that defined contribution plans may provide a strong incentive for insurance plans to lower costs. However, because there is substantial heterogeneity in price sensitivity, the health insurance marketplace may be comprised of a spectrum of consumers with different propensities to “shop around.” New enrollees are more sensitive to prices and also more likely to respond differentially according to health. Thus, although we see less adverse selection among old enrollees, the adverse selection that does occur with new enrollees can become entrenched once these consumers choose a plan and are then old enrollees with respect to any future changes in price. Price competition following these dynamics may be limited, and insurance providers who gain an early share of the healthy population may be able to extract rents from the reluctance of enrollees to switch plans.

We intend for this study to provide some initial estimates of the price responsiveness of low-income families choosing health insurance, and we note potentially useful extensions to this study. First, we take prices paid by enrollees as exogenous. Unobserved characteristics may likely increase both prices and demand (as suggested by the positive coefficient on previous-year price), and price sensitivity may be even greater than the already substantial estimates that we find. There are a number of features of this insurance “marketplace” that suggest that at least some part of the prices that consumers face are exogenous and may be used as instruments to refine estimates of consumer responses to exogenous price changes. For example, base contribution rates are determined by the state Commonwealth Connector Authority and are linked to percentages of median income levels; firm price bids are made according to statewide demographic groups, which are then translated by demographic formulas to prices in local marketplaces in which enrollees would be making decisions.

Secondly, our analysis is limited to new- and old-enrollee choices in 2008 and may be extended for a more global understanding of decision-making. For new enrollees, we study the decision of choosing a plan, conditional on enrolling, and have therefore described an intensive margin of plan choice rather than an extensive one that includes the possibility of no enrollment. Although the number of uninsured in Massachusetts has declined to low levels, due partly to affordable plans and to a state mandate, price responsiveness along both intensive and extensive margins would necessarily be even greater than responsiveness along the intensive margin alone. For old enrollees, we focus on the single choice of whether to stay or switch to another plan in 2008, although we account for previous-year plan in this choice. An extension might exploit the panel structure of the data and also model the original choice to enroll in a plan in 2007, thereby considering both choices as sequential decisions. This approach may be useful to characterize the dynamics of plan choice, particularly with respect to adverse selection, and may be used to compare new enrollees in 2007 with new enrollees in 2008.

Table 3.1.1: Changes in Enrollee Contribution by Plan Type

Plan Type	Change in cost		Cost range in 2007		Cost range in 2008	
	Mean	SD	Mean	SD	Mean	SD
IIA	\$9.14	\$13.78	\$0.00	\$0.00	\$24.15	\$10.04
IIB	\$14.04	\$25.63	\$40.19	\$14.49	\$47.62	\$20.48
IIIA	\$48.66	\$39.95	\$47.72	\$24.11	\$56.99	\$29.89
IIIB	\$50.85	\$40.46	\$48.96	\$33.09	\$57.10	\$29.79
IVA	\$16.08	\$42.15	\$38.74	\$20.80	\$57.92	\$30.48
IVB	\$21.47	\$39.27	\$33.65	\$16.56	\$55.49	\$30.15
Total	\$18.83	\$30.22	\$24.52	\$26.07	\$39.93	\$24.36

Note: Numbers in the first set of columns represent the means and standard deviations of the change in enrollee contribution for each plan averaged across areas, plan type, and insurer. The next two sets of columns represent the difference in cost between the most expensive and cheapest plans for each area and plan type in 2007 and 2008. Enrollee contributions changed from "2007" to "2008" in July 2008. Plan types are as follows: IIA for those with incomes 100-150% of poverty, IIB for those 150-200% of poverty, IIIA or IVA for those 200-250% of poverty, and IIIB or IVB those 250-300% of poverty in plan type IIIB. Plan types IVA and IVB corresponded to lower-copayment, higher-premium options that were discontinued in 2008. To calculate the change in cost and 2008 cost range for IVA and IVB, we assume that enrollees continued in the corresponding plan by the same insurer and in the same area in IIIA and IIIB, respectively.

Table 3.2.1: Switches in Existing Enrollment from June 2008 to September 2008

Plan Type	June 2008					September 2008				
	MMCO					MMCO				
	BMC	Fallon	NHP	Network	Total	BMC	Fallon	NHP	Network	Total
IIA	13,429	1,284	5,835	11,238	31,786	14,568	1,188	4,977	11,053	31,786
IIB	11,996	486	2,128	10,881	25,491	13,544	486	2,250	9,211	25,491
IIIA	4,716	420	788	3,750	9,674	4,287	1879	2966	2,813	11,945
IIIB	2,121	236	407	1,722	4,486	2,291	892	1312	1,467	5,962
IVA	1,157	130	368	616	2271					
IVB	785	89	179	423	1476					
Total	34,204	2,645	9,705	28,630	75,184	34,690	4,445	11,505	24,544	75,184

Note: Numbers represent enrollees by June 2008 and the same enrollees who may have switched plans by September 2008 ("old enrollees"). Enrollees with incomes 100-150% of poverty were in plan type IIA, those 150-200% of poverty in plan type IIB, those 200-250% of poverty in plan type IIIA (lower premium but higher copayment) or plan type IVA (higher premium but lower copayment), those 250-300% of poverty in plan type IIIB (lower premium but higher copayment) or plan type IVB (higher premium but lower copayment). Enrollee contributions changed from "2007" to "2008" in July 2008. Plan types IVA and IVB no longer were available in 2008. MMCOs include Boston Medical Center HealthNet Plan (BMC), Fallon Community Health Plan (Fallon), Neighborhood Health Plan (NHP), and Network Health.

Table 3.2.2: New Enrollment in 2008 by Plan Type and MMCO

Plan Type	MMCO				Total
	BMC	Fallon	NHP	Network	
I	24,392	3,616	13,283	21,412	62,703
IIA	8,775	892	3,068	5,207	17,942
IIB	9,555	433	1,518	7,873	19,379
IIIA	4,046	983	1,738	3,161	9,928
IIIB	2,107	503	926	1,522	5,058
Total	48,875	6,427	20,533	39,175	115,010

Note: Numbers represent numbers of new enrollees in each plan type for each Medicaid Managed Care Organization (MMCO) in 2008. MMCOs include Boston Medical Center HealthNet Plan (BMC), Fallon Community Health Plan (Fallon), Neighborhood Health Plan (NHP), and Network Health. Plan types are as defined above: IIA for those with incomes 100-150% of poverty, IIB for 150-200% of poverty, IIIA (low premium, high copayment) and IVA (high premium, low copayment) for 200-250% of poverty, and IIIB (low premium, high copayment) and IVB (high premium, low copayment) for 250-300% of poverty. In addition, plan type I is for enrollees below poverty.

Table 3.2.3: New Enrollment in 2008 by Month and Plan Type

Month	Plan Type					Total
	I	IIA	IIB	IIIA	IIIB	
January	6,583	1,674	3,491	1,854	945	14,547
February	6,935	2,659	2,645	1,486	748	14,473
March	5,787	1,845	2,294	1,148	557	11,631
April	5,340	1,653	1,862	874	411	10,140
May	4,874	1,550	1,530	708	405	9,067
June	5,339	1,455	1,168	650	313	8,925
July	4,627	1,311	1,266	664	303	8,171
August	4,780	1,138	1,029	434	261	7,642
September	4,575	1,138	975	521	296	7,505
October	4,697	1,206	1,000	506	263	7,672
November	4,992	1,206	905	462	223	7,788
Total	62,703	17,942	19,379	9,928	5,058	115,010

Note: Numbers represent numbers of new enrollees in each plan type and each month in 2008. MMCOs include Boston Medical Center HealthNet Plan (BMC), Fallon Community Health Plan (Fallon), Neighborhood Health Plan (NHP), and Network Health. Plan types are as defined above: IIA for those with incomes 100-150% of poverty, IIB for 150-200% of poverty, IIIA (low premium, high copayment) and IVA (high premium, low copayment) for 200-250% of poverty, and IIIB (low premium, high copayment) and IVB (high premium, low copayment) for 250-300% of poverty. In addition, plan type I is for those below poverty.

Table 3.3.1: Plan Choice Conditional Logit Regressions

	Old Enrollee Choice				New Enrollee Choice	
	(1)	(2)	(3)	(4)	(5)	(6)
Previous Price (2007)		0.027* (0.006)	0.023* (0.005)	0.023* (0.004)		
Price (2008)	-0.045* (0.002)	-0.044* (0.002)	-0.076* (0.006)	-0.078* (0.006)	-0.027* (0.002)	-0.044* (0.005)
Price × Q2			0.003* (0.001)	0.001 (0.001)		0.004* (0.001)
Price × Q3			0.001 (0.001)	0.005* (0.001)		0.012* (0.002)
Price × Q4			0.003* (0.001)	0.009* (0.001)		0.013* (0.001)
Price × IIB			0.021* (0.004)	0.020* (0.004)		0.009* (0.003)
Price × IIIA			0.031* (0.009)	0.031* (0.009)		0.009 (0.005)
Price × IIIB			0.035* (0.009)	0.035* (0.009)		0.009 (0.005)
Price × IVA			0.046* (0.008)	0.045* (0.008)		
Price × IVB			0.046* (0.009)	0.045* (0.009)		
Fallon	-0.317* (0.128)	-0.685* (0.149)	0.069 (0.289)	0.053 (0.289)	-0.793* (0.223)	-0.764* (0.242)
NHP	0.162 (0.127)	-0.639* (0.163)	0.016 (0.203)	0.008 (0.203)	-0.565* (0.178)	-0.541* (0.179)
Network Health	-0.064 (0.096)	-0.101 (0.092)	0.189 (0.142)	0.187 (0.140)	-0.012 (0.199)	-0.005 (0.202)

Note: Model coefficients with standard errors in parentheses are presented. For old enrollees (those enrolled before July 2008), "cost" is taken to be 2008 cost, while "previous cost" is the 2007 cost of the same alternative. "Cost" for new enrollees in model (6) is slightly different, in that 2007 cost figures are used for enrollees before July. Cost is interacted with dummies for health risk quartile and plan type in models (3), (4), and (6), where health risk quartile is denoted by "Q". Different definitions of health risk quartile are used: demographic-based health risk is used for models (3) and (6), while health care utilization is used in model (4), which is only available for old enrollees. Plan types are as defined above: IIA for those with incomes 100-150% of poverty, IIB for 150-200% of poverty, IIIA (low premium, high copayment) and IVA (high premium, low copayment) for 200-250% of poverty, and IIIB (low premium, high copayment) and IVB (high premium, low copayment) for 250-300% of poverty. Plan type I is not used, since these plans are free. Interaction terms are represented more conveniently in Table 3.3.5. * Denotes significance at the 5% level.

Table 3.3.2: Percent Change in Demand with \$10 Own- and Cross-price Increase

MMCO Impacted	Old Enrollees			
	MMCO with Price Change			
	BMC	Fallon	NHP	Network
BMC	-7.2%	0.5%	3.1%	4.3%
Fallon	3.0%	-16.1%	4.8%	8.7%
NHP	7.4%	1.9%	-16.2%	6.3%
Network	4.7%	1.4%	3.0%	-9.7%

MMCO Impacted	New Enrollees			
	MMCO with Price Change			
	BMC	Fallon	NHP	Network
BMC	-15.3%	1.1%	3.8%	10.1%
Fallon	8.8%	-20.9%	5.2%	7.8%
NHP	12.0%	2.0%	-21.7%	9.2%
Network	12.1%	1.2%	3.5%	-17.1%

Note: Percentages represent the relative percent change in enrollment (aggregated across plan types and areas) for old enrollees and new enrollees by panel. Columns represent the MMCO that undergoes a \$10 increase in price, while rows represent the MMCO that is impacted by the price change. MMCOs are Boston Medical Center HealthNet Plan (BMC), Fallon Community Health Plan (Fallon), Neighborhood Health Plan (NHP), and Network Health.

Table 3.3.3: Own- and Cross-price Elasticities of Demand

MMCO Impacted	Old Enrollees			
	MMCO with Price Change			
	BMC	Fallon	NHP	Network
BMC	-0.349	0.032	0.197	0.236
Fallon	0.224	-1.249	0.452	1.055
NHP	0.512	0.149	-1.099	0.447
Network	0.142	0.117	0.195	-0.687

MMCO Impacted	New Enrollees			
	MMCO with Price Change			
	BMC	Fallon	NHP	Network
BMC	-0.548	0.063	0.195	0.370
Fallon	0.387	-1.228	0.376	0.387
NHP	0.406	0.125	-1.171	0.345
Network	0.431	0.067	0.184	-0.646

Note: Numbers are own-price and cross-price elasticities of demand (aggregated across plan types and areas) for old enrollees and new enrollees by panel. Columns represent the MMCO that undergoes a price change, while rows represent the MMCO that is impacted by the price change. MMCOs are Boston Medical Center HealthNet Plan (BMC), Fallon Community Health Plan (Fallon), Neighborhood Health Plan (NHP), and Network Health.

Table 3.3.4: Percent Change in Demand with \$10 Own-price Increase by Subgroup of Enrollees

Enrollee Population (Model in Parentheses)							
Enrollee Subgroup	Old Enrollees						New Enrollees
	All (3)	All (4)	Lowest (4)	Highest (4)	Stayed (4)	Switched (4)	All (6)
Overall	-10.1% (9.7%)	-10.1% (9.7%)	-11.7% (10.6%)	-8.1% (8.5%)	-8.5% (8.9%)	-20.5% (7.9%)	-17.2% (5.2%)
Q1	-10.3% (9.7%)	-11.1% (10.5%)	-13.4% (11.5%)	-10.0% (10.1%)	-9.1% (9.6%)	-22.3% (7.8%)	-21.2% (4.5%)
Q2	-9.4% (8.9%)	-11.1% (10.6%)	-12.9% (11.2%)	-8.7% (9.1%)	-9.1% (9.7%)	-22.3% (7.8%)	-19.6% (5.4%)
Q3	-10.5% (10.0%)	-9.9% (9.4%)	-11.0% (10.2%)	-8.0% (8.3%)	-8.3% (8.6%)	-20.1% (7.6%)	-15.0% (3.9%)
Q4	-10.0% (9.9%)	-8.3% (7.8%)	-9.5% (8.6%)	-6.5% (6.5%)	-7.3% (7.2%)	-16.4% (6.9%)	-14.0% (2.8%)
IIA	-9.8% (9.6%)	-9.8% (9.5%)			-8.9% (9.0%)	-20.3% (9.0%)	-21.9% (4.6%)
IIB	-7.8% (9.7%)	-7.7% (9.6%)	-8.7% (10.9%)	-14.8% (6.6%)	-6.1% (8.1%)	-23.3% (9.1%)	-14.7% (3.2%)
IIIA	-16.6% (8.4%)	-16.7% (8.7%)	-19.7% (6.6%)	-1.2% (1.3%)	-14.3% (9.2%)	-21.0% (5.5%)	-15.1% (3.6%)
IIIB	-14.0% (7.9%)	-14.1% (8.3%)	-17.4% (6.5%)	-4.4% (7.8%)	-12.1% (8.4%)	-19.1% (5.7%)	-14.5% (4.0%)
IVA	-6.9% (5.0%)	-7.0% (4.9%)	-8.7% (3.2%)	-0.7% (0.6%)	-6.3% (4.6%)	-10.7% (4.3%)	
IVB	-7.0% (5.3%)	-7.1% (5.0%)	-8.8% (3.7%)	-0.8% (0.6%)	-6.4% (4.9%)	-10.4% (4.6%)	

Note: Numbers represent mean own-price percent changes in demand, averaged over the relevant population and weighted by the likelihood of each enrollee to pick a given plan. Standard deviations are given in parentheses. Rows represent subgroups by health-risk quartile (based on demographic or claims-based data, depending on the model used). Columns represent the overall population, including old enrollees whose elasticities represent switching decisions and new enrollees in 2008 whose elasticities represent new enrollment decisions. In particular, the columns labeled "Lowest" and "Highest" refer to old enrollees previously in the cheapest or most expensive plan, respectively. Columns labeled "Stayed" and "Switched" represent old enrollees who stayed in or switched from their previous plan, respectively. Note that plan types IVA and IVB were not available for new enrollees. Also, all plans in IIA were previously free, so that the "Lowest" and "Highest" columns are not relevant for this plan type.

Table 3.3.5: Own-price Elasticities of Demand by Subgroup of Enrollees

Enrollee Population (Model in Parentheses)							
Enrollee Subgroup	Old Enrollees						New Enrollees
	All (3)	All (4)	Lowest (4)	Highest (4)	Stayed (4)	Switched (4)	All (6)
Overall	-0.647 (0.984)	-0.647 (0.992)	-1.174 (1.219)	-0.608 (0.727)	-0.467 (0.826)	-1.799 (1.179)	-0.714 (0.655)
Q1	-0.621 (1.012)	-0.711 (1.109)	-1.357 (1.348)	-0.793 (0.962)	-0.492 (0.913)	-1.979 (1.279)	-0.674 (0.748)
Q2	-0.642 (0.977)	-0.727 (1.093)	-1.304 (1.312)	-0.648 (0.782)	-0.508 (0.903)	-1.969 (1.235)	-0.799 (0.714)
Q3	-0.696 (1.003)	-0.625 (0.935)	-1.095 (1.154)	-0.585 (0.651)	-0.458 (0.784)	-1.743 (1.086)	-0.667 (0.580)
Q4	-0.633 (0.926)	-0.527 (0.781)	-0.935 (0.973)	-0.482 (0.522)	-0.416 (0.683)	-1.401 (0.935)	-0.679 (0.570)
IIA	-0.125 (0.185)	-0.124 (0.183)			-0.108 (0.174)	-0.317 (0.177)	-0.055 (0.103)
IIB	-0.545 (0.827)	-0.540 (0.823)	-0.642 (0.945)	-0.984 (0.493)	-0.399 (0.684)	-1.893 (0.810)	-0.687 (0.137)
IIIA	-1.863 (1.061)	-1.866 (1.086)	-2.211 (0.886)	-0.111 (0.115)	-1.573 (1.112)	-2.400 (0.796)	-1.366 (0.390)
IIIB	-2.148 (1.332)	-2.164 (1.387)	-2.692 (1.148)	-0.711 (1.314)	-1.819 (1.366)	-2.998 (1.041)	-1.868 (0.568)
IVA	-0.801 (0.657)	-0.808 (0.642)	-0.969 (0.488)	-0.069 (0.052)	-0.713 (0.601)	-1.302 (0.625)	
IVB	-1.082 (0.883)	-1.095 (0.850)	-1.318 (0.665)	-0.106 (0.083)	-0.989 (0.815)	-1.664 (0.810)	

Note: Numbers represent mean own-price elasticities, averaged over the relevant population and weighted by the likelihood of each enrollee to pick a given plan. Standard deviations are given in parentheses. Rows represent subgroups by health-risk quartile (based on demographic or claims-based data, depending on the model used). Columns represent the overall population, including old enrollees whose elasticities represent switching decisions and new enrollees in 2008 whose elasticities represent new enrollment decisions. In particular, the columns labeled "Lowest" and "Highest" refer to old enrollees previously in the cheapest or most expensive plan, respectively. Columns labeled "Stayed" and "Switched" represent old enrollees who stayed in or switched from their previous plan, respectively. Note that plan types IVA and IVB were not available for new enrollees. Also, all plans in IIA were previously free, so that the "Lowest" and "Highest" columns are not relevant for this plan type.

Bibliography

- ABADIE, A. (2005): "Semiparametric difference-in-differences estimators," *Review of Economic Studies*, 72, 1–19.
- (2010): "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 105, 493–505.
- ABADIE, A., A. DIAMOND, AND J. HAINMUELLER (2011): "Comparative politics and the synthetic control method," *MIT Political Science Department Research Paper*, No. 2011-25.
- AGHION, P. AND J. TIROLE (1997): "Formal and real authority in organizations," *The Journal of Political Economy*, 105, 1–29.
- ANDERSON, G. F., P. S. HUSSEY, B. K. FROGNER, AND H. R. WATERS (2005): "Health spending in the United States and the rest of the industrialized world," *Health Affairs*, 24, 903–914.
- BANDIERA, O., I. BARANKAY, AND I. RASUL (2005): "Social preferences and the response to incentives: Evidence from personnel data," *The Quarterly Journal of Economics*, 120, 917–962.
- (2009): "Social connections and incentives in the workplace: Evidence from personnel data," *Econometrica*, 77, 1047–1094.
- (2010): "Social incentives in the workplace," *The Review of Economic Studies*, 77, 417–458.
- BANERJEE, A. V. (1992): "A simple model of herd behavior," *The Quarterly Journal of Economics*, 107, 797–817.
- BATES, D. W., G. J. KUPERMAN, E. RITTENBERG, J. M. TEICH, J. FISKIO, N. MA'LUF, A. ONDERDONK, D. WYBENGA, J. WINKELMAN, AND T. A. BRENNAN (1999): "A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests," *The American Journal of Medicine*, 106, 144–150.
- BERNSTEIN, S. L., D. ARONSKY, R. DUSEJA, S. EPSTEIN, D. HANDEL, U. HWANG, M. MCCARTHY, K. JOHN MCCONNELL, J. M. PINES, AND N. RATHLEV (2008): "The effect of emer-

- gency department crowding on clinically oriented outcomes," *Academic Emergency Medicine*, 16, 1–10.
- BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, 119, 249–275.
- BLISS, C. AND B. NALEBUFF (1984): "Dragon-slaying and ballroom dancing: The private supply of a public good," *Journal of Public Economics*, 25, 1–12.
- BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2011): "Does management matter? Evidence from India," Tech. Rep. 16658, National Bureau of Economic Research.
- BLOOM, N. AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices Across Firms and Countries," *The Quarterly Journal of Economics*, 122, 1351–1408.
- BOHNET, I. AND S. HUCK (2004): "Repetition and reputation: Implications for trust and trustworthiness when institutions change," *The American Economic Review*, 94, 362–366.
- BURKE, M. A., G. M. FOURNIER, AND K. PRASAD (2010): "Geographic variations in a model of physician treatment choice with social interactions," *Journal of Economic Behavior and Organization*, 73, 418–432.
- CHANDRA, A. AND D. O. STAIGER (2007): "Productivity spillovers in healthcare: evidence from the treatment of heart attacks," *The Journal of Political Economy*, 115, 103–140.
- CHE, Y. K. AND N. KARTIK (2009): "Opinions as incentives," *The Journal of Political Economy*, 117, 815–860.
- CHOUHRY, N. K., G. M. ANDERSON, A. LAUPACIS, D. ROSS-DEGNAN, S. L. T. NORMAND, AND S. B. SOUMERAI (2006): "Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis," *British Medical Journal*, 332, 141–145.
- COLEMAN, J. S., E. KATZ, AND H. MENZEL (1966): *Medical Innovation: A Diffusion Study*, Bobbs-Merrill Company New York, NY.
- CRAWFORD, V. P. AND J. SOBEL (1982): "Strategic information transmission," *Econometrica*, 50, 1431–1451.
- CUTLER, D. AND S. REBER (1998): "Paying for health insurance: The trade-off between competition and adverse selection," *The Quarterly Journal of Economics*, 113, 433–466.
- ELLISON, G. AND D. FUDENBERG (1993): "Rules of thumb for social learning," *Journal of Political Economy*, 101, 612–643.
- EPSTEIN, A. J. AND S. NICHOLSON (2009): "The formation and evolution of physician treatment styles: an application to cesarean sections," *Journal of Health Economics*, 28, 1126–1140.

- FEHR, E. AND S. GACHTER (2000): "Cooperation and punishment in public goods experiments," *The American Economic Review*, 90, 980–994.
- (2002): "Altruistic punishment in humans," *Nature*, 415, 137–140.
- FEHR, E. AND K. M. SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *The Quarterly Journal of Economics*, 114, 817–868.
- FISHER, E. S., D. E. WENNBURG, T. A. STUKEL, D. J. GOTTLIEB, AND F. L. LUCAS (2003a): "The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care," *Annals of Internal Medicine*, 138, 288–298.
- FISHER, E. S., D. E. WENNBURG, T. A. STUKEL, D. J. GOTTLIEB, F. L. LUCAS, AND E. L. PINDER (2003b): "The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of care," *Annals of Internal Medicine*, 138, 273–287.
- FORSTER, A. J., I. STIELL, G. WELLS, A. J. LEE, AND C. VAN WALRAVEN (2003): "The effect of hospital occupancy on emergency department length of stay and patient disposition," *Academic Emergency Medicine*, 10, 127–133.
- FUDENBERG, D. AND D. K. LEVINE (1993): "Self-confirming equilibrium," *Econometrica*, 61, 523–545.
- GAWANDE, A. (2009): "The cost conundrum," *The New Yorker*, June, 36–44.
- (2012): "Big Med: Restaurant chains have managed to combine quality control, cost control, and innovation. Can health care?" *The New Yorker*.
- GIBBONS, R. AND R. HENDERSON (2012): "What do managers do? Exploring persistent performance differences among seemingly similar enterprises," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton, NJ: Princeton University Press, 680–732.
- GREIF, A. (1993): "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," *The American Economic Review*, 83, 525–548.
- GRYTEN, J. AND R. SORENSEN (2003): "Practice variation and physician-specific effects," *Journal of Health Economics*, 22, 403–418.
- HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2003): "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation," *The Journal of Political Economy*, 111, 465–497.
- HOLMSTROM, B. (1982): "Moral hazard in teams," *The Bell Journal of Economics*, 13, 324–340.

- ICHNIOWSKI, C., K. SHAW, AND G. PRENNUSHI (1997): "The effects of human resource management practices on productivity: A study of steel finishing lines," *The American Economic Review*, 87, 291–313.
- INSTITUTE OF MEDICINE (2012): "Best care at lower cost: The path to continuously learning health care in America," Tech. rep., National Academies Press, Washington, DC.
- IWASHYNA, T. J., A. FULD, D. A. ASCH, AND L. M. BELLINI (2011): "The impact of residents, interns, and attendings on inpatient laboratory ordering patterns: A report from one university's hospitalist service," *Academic Medicine*, 86, 139.
- JACKSON, C. K. AND H. S. SCHNEIDER (2010): "Do social connections reduce moral hazard? Evidence from the New York City taxi industry," Tech. Rep. 16279, National Bureau of Economic Research.
- KAMENICA, E. (2008): "Contextual inference in markets: On the informational content of product lines," *The American Economic Review*, 98, 2127–2149.
- KANDEL, E. AND E. P. LAZEAR (1992): "Peer pressure and partnerships," *The Journal of Political Economy*, 801–817.
- LEE, T. H. AND J. J. MONGAN (2009): *Chaos and Organization in Health Care*, Cambridge, MA: MIT Press.
- LERMAN, B. AND M. S. KOBERNICK (1987): "Return visits to the emergency department," *The Journal of Emergency Medicine*, 5, 359–362.
- LIBERMAN, V., S. M. SAMUELS, AND L. ROSS (2004): "The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves," *Personality and Social Psychology Bulletin*, 30, 1175–1185.
- MANSKI, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60, 531–542.
- MAS, A. AND E. MORETTI (2009): "Peers at work," *The American Economic Review*, 99, 112–145.
- MCCARTHY, D. AND D. BLUMENTHAL (2006): "Stories from the sharp end: case studies in safety improvement," *Milbank Quarterly*, 84, 165–200.
- MCCARTHY, D. AND K. MUELLER (2009): "Organizing for higher performance: Case studies of organized delivery systems," Tech. Rep. 1288, The Commonwealth Fund.
- MC HUGH, M., K. VAN DYKE, M. MCCLELLAND, AND D. MOSS (2011): "Improving patient flow and reducing emergency department crowding: A guide for hospitals," Tech. Rep. 11(12)-0094, Agency for Healthcare Research and Quality.

- MILGROM, P. AND J. ROBERTS (1988): "An economic approach to influence activities in organizations," *The American Journal of Sociology*, 94, 154–179.
- MOLITOR, D. (2011): "The evolution of physician practice styles: Evidence from cardiologist migration," Tech. rep., Massachusetts Institute of Technology.
- OLIVER, A. (2007): "The Veterans Health Administration: an American success story?" *Milbank Quarterly*, 85, 5–35.
- OSTROM, E., J. WALKER, AND R. GARDNER (1992): "Covenants with and without a sword: Self-governance is possible," *The American Political Science Review*, 404–417.
- OTTAVIANI, M. AND P. SORENSEN (2001): "Information aggregation in debate: who should speak first?" *Journal of Public Economics*, 81, 393–421.
- PALLAK, M. S. AND R. O. PERLOFF (1986): *Psychology and Work: Productivity, Change, and Employment.*, Washington DC: American Psychological Association.
- PATEL, P. B. AND D. R. VINSON (2005): "Team assignment system: Expediting emergency department care." *Annals of Emergency Medicine*, 46, 499.
- PHELPS, C. E. (2000): "Information diffusion and best practice adoption," in *Handbook of Health Economics*, Amsterdam: Elsevier, 223–264.
- PHELPS, C. E. AND C. MOONEY (1993): "Variations in medical practice use: causes and consequences," *Competitive Approaches to Health Care Reform*, 139–175.
- RANTAKARI, H. (2011): "Organizational design and environmental volatility," *Journal of Law, Economics, and Organization*, doi:10.1093/jleo/ewr015.
- ROSENBAUM, P. (2002): "Covariance adjustment in randomized experiments and observational studies," *Statistical Science*, 17, 286–304.
- RUBINSTEIN, A. (1982): "Perfect equilibrium in a bargaining model," *Econometrica*, 50, 97–109.
- SALLUZZO, R. F., T. A. MAYER, R. W. STRAUSS, AND P. KIDD (1997): *Emergency Department Management: Principles and Application*, Mosby St Louis, MO.
- SCHARFSTEIN, D. S. AND J. C. STEIN (1990): "Herd behavior and investment," *The American Economic Review*, 80, 465–479.
- SCHROEDER, S. A., A. SCHLIFTMAN, AND T. E. PIEMME (1974): "Variation among physicians in use of laboratory tests: relation to quality of care," *Medical Care*, 12, 709–713.
- SCHUUR, J. D. AND A. K. VENKATESH (2012): "The growing role of emergency departments in hospital admissions," *The New England Journal of Medicine*, 367, 391–393.

- SEARLE, S. R., G. CASELLA, AND C. E. MCCULLOCH (1992): *Variance Components*, Wiley New York.
- THOMPSON, D. A., P. R. YARNOLD, D. R. WILLIAMS, AND S. L. ADAMS (1996): "Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department." *Annals of Emergency Medicine*, 28, 657–665.
- VAN DEN STEEN, E. (2010): "On the origin of shared beliefs (and corporate culture)," *The RAND Journal of Economics*, 41, 617–648.
- WEBER, R. A. AND C. F. CAMERER (2003): "Cultural conflict and merger failure: An experimental approach," *Management Science*, 49, 400–415.
- WENNBERG, J. AND A. GITTELSON (1973): "Small area variations in health care delivery," *Science*, 182, 1102–1108.
- WENNBERG, J. E., E. S. FISHER, AND J. S. SKINNER (2002): "Geography and the debate over Medicare reform," *Health Affairs*, 21, 10–10.
- WENNBERG, J. E., E. S. FISHER, T. A. STUKEL, AND S. M. SHARP (2004): "Use of Medicare claims data to monitor provider-specific performance among patients with severe chronic illness," *Health Affairs*, 23, 5–18.
- YEATTS, D. E. AND C. HYTEN (1997): *High-performing self-managed work teams: A comparison of theory to practice*, Sage Publications, Incorporated.